

ENERGY-EFFICIENT CIRCUIT DESIGN

by

Michael Nugent

B.S. Computer Science and Mathematics, University of Pittsburgh,

2007

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial

fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH
DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Michael Nugent

It was defended on

April 29, 2015

and approved by

Kirk Pruhs, Department of Computer Science

Daniel Mossé, Department of Computer Science

Adam Lee, Department of Computer Science

Anupam Gupta, Department of Computer Science, Carnegie Mellon University

Dissertation Director: Kirk Pruhs, Department of Computer Science

ENERGY-EFFICIENT CIRCUIT DESIGN

Michael Nugent, PhD

University of Pittsburgh, 2015

We initiate the theoretical investigation of energy-efficient circuit design. We assume that the circuit design specifies the circuit layout as well as the supply voltages for the gates. To obtain maximum energy efficiency, the circuit design must balance the conflicting demands of minimizing the energy used per gate, and minimizing the number of gates in the circuit; If the energy supplied to the gates is small, then functional failures are likely, necessitating a circuit layout that is more fault-tolerant, and thus that has more gates.

By leveraging previous work on fault-tolerant circuit design, we show general upper and lower bounds on the amount of energy required by a circuit to compute a given relation. We show that some circuits would be asymptotically more energy-efficient if heterogeneous supply voltages were allowed, and show that for some circuits the most energy-efficient supply voltages are homogeneous over all gates.

In the traditional approach to circuit design the supply voltages for each transistor/gate are set sufficiently high so that with sufficiently high probability no transistor fails. We show that if there is a better (in terms of worst-case relative error with respect to energy) method than the traditional approach then $P = NP$, and thus there is a complexity theoretic obstacle to achieving energy savings with Near-Threshold computing.

We show that almost all Boolean functions require circuits that use exponential energy. This is not an immediate consequence of Shannon's classic result that most functions require exponential sized circuits of faultless gates because, as we show, the same circuit layout can compute many different functions, depending on the value of the supply voltage.

If the error bound must vanish as the number of inputs increases, we show that a natural

class of functions can be computed with asymptotically less energy using heterogeneous supply voltages than is possible using homogeneous supply voltages. We also prove upper bounds on the asymptotic energy savings achieved by using heterogeneous supply voltages over homogeneous supply voltages for a class of functions, and also show a relation that can bypass this bound.

TABLE OF CONTENTS

PREFACE	viii
1.0 INTRODUCTION	1
1.1 Related Work	5
1.2 Our Contributions	7
1.2.1 General Bounds on Circuits With Constant Error	7
1.2.2 Introduction to Heterogeneity	8
1.2.3 Hardness Results	10
1.2.4 Almost All Functions Require Exponential Energy	13
1.2.5 The Power of Heterogeneity to Reduce Energy	15
2.0 MODEL, DEFINITIONS, AND NOTATION	19
3.0 GENERAL ENERGY UPPER AND LOWER BOUNDS	21
3.1 A General Energy Lower Bound	21
3.2 A General Energy Upper Bound	28
4.0 INTRODUCTION TO SUPPLY VOLTAGE HETEROGENEITY	29
4.1 Supply Voltage Heterogeneity May Not Help	29
4.2 Supply Voltage Heterogeneity Can Help	30
5.0 HARDNESS AND ALGORITHMIC RESULTS FOR CIRCUIT ENERGY PROBLEMS	38
5.1 Polynomial-Time Approximation of the Minimum Circuit Energy Problem	39
5.2 Hardness of Approximation for the Minimum Circuit Energy Problem	42
5.3 Hardness of Determining (ϵ, δ) -reliability on Fixed Inputs	48
5.4 Tree Circuits	53

5.5	Non-Monotonicity of δ in ϵ	54
6.0	ALMOST ALL FUNCTIONS REQUIRE EXPONENTIAL ENERGY	57
6.1	A Lower Bound on the Number of Functions Computable by a Circuit	57
6.1.1	Homogeneous Supply Voltages	58
6.1.2	Heterogeneous Supply Voltages	62
6.2	Almost all Functions Require Exponential Energy	65
6.2.1	Adaptation of Shannon's Argument	66
6.2.2	Homogeneous Supply Voltages	66
6.2.3	Heterogeneous Supply Voltages	68
6.3	Relating Energy and the Number of Noisy Gates	70
7.0	THE POWER OF HETEROGENEITY TO REDUCE ENERGY	72
7.1	Lower Bound for Functions	72
7.2	Upper Bound for Functions	74
7.3	Lower Bound for Relations	76
7.4	Generalizing the Failure-to-Energy Function	81
8.0	CONCLUSION	83
8.1	Open Problems	84
8.1.1	Solving the Minimum Circuit Energy Problem for Restricted Classes of Circuits	85
8.1.2	Whether Heterogeneity Reduces Energy When δ is a Fixed Constant	85
8.1.3	Whether $\log n$ Energy Savings via Heterogeneity is the Maximum Pos- sible When δ Vanishes	86
8.1.4	The Power of the Exact Failure Model	86
	BIBLIOGRAPHY	88

LIST OF FIGURES

1	Semi-log plot of voltage-to-failure for an SRAM cell from [16].	2
2	Two SRAM circuits with the same functionality.	3
3	(a) $\Pr[r \text{ outputs } 1] \geq 1 - p$. (b) The path from b to g . The input gates b_i receive input 0.	34
4	A subtree B . The solid edges denote the full ternary subtree $T \in \Gamma$. Note that T has 1's as inputs on its leafs. The dashed edges denote the edges in $B \setminus T$. The gray nodes denote gates that failed.	36
5	The circuit S_ϕ where $\phi = (x_1 \vee \bar{x}_2 \vee x_4) \wedge (x_1 \vee x_2 \vee x_3) \wedge (\bar{x}_3 \vee x_6 \vee x_5) \wedge (x_3 \vee \bar{x}_5 \vee \bar{x}_6)$	43
6	A simple circuit where $\delta^*(\epsilon)$ is not monotone in ϵ in the von Neumann failure model, consisting of two OR gates and one AND gate.	55
7	A simple circuit where $\delta^*(\epsilon)$ is not monotone in ϵ in the 0-default failure model, consisting of an AND gate and two NOT gates.	56
8	The circuit used in the proof of Theorem 53.	59
9	A tree of adders. The majority of x_1, \dots, x_n is a function of $y_1, \dots, y_{\log n}$ that is computable by a circuit of size $o(n)$	78
10	A 1-bit full adder: logic block (left) and circuit realization (right).	79

PREFACE

The research within this dissertation was published in the following papers, and was a collaboration with the listed coauthors:

- *Energy-Efficient Circuit Design*, published in the 5th conference on Innovations in Theoretical Computer Science (ITCS), with Antonios Antoniadis, Neal Barcelo, Kirk Pruhs, and Michele Scquizzato [6].
- *Complexity-Theoretic Obstacles to Achieving Energy Savings with Near-Threshold Computing*, published in the 5th International Green Computing Conference (IGCC), with Antonios Antoniadis, Neal Barcelo, Kirk Pruhs, and Michele Scquizzato [5].
- *Almost All Functions Require Exponential Energy*, to appear in the 40th International Symposium on Mathematical Foundations of Computer Science (MFCS), with Neal Barcelo, Kirk Pruhs, and Michele Scquizzato [8].
- *The Power of Heterogeneity in Near-Threshold Computing*, in submission to the 6th International Green Computing Conference (IGCC), with Neal Barcelo, Kirk Pruhs, and Michele Scquizzato [9].

We thank Rami Melhem for insightful discussions about Near-Threshold Computing.

1.0 INTRODUCTION

The number of transistors per unit volume on a chip continues to double about every two years. However, about a decade ago chip makers hit a thermal wall as the cost of cooling chips with these transistor densities became prohibitive. This has resulted in Moore’s gap, namely that increased transistor density no longer directly translates into a similar increase in performance, and in energy becoming the first order design constraint in CMOS-based technologies.

One possible technique to attain more energy-efficient circuits is *Near-Threshold Computing*. The threshold voltage of a transistor is the minimum voltage at which the transistor starts to conduct current, around 0.2-0.3V for modern processors. Of course, even for identically-designed transistors, there can be variations in the actual threshold voltage due to manufacturing variations; And even for the same transistor, the actual threshold voltage will vary with environmental conditions. Further, actual supply voltages may differ from the designed voltage due to manufacturing and environmental variance. Thus if the designed supply voltage was exactly the ideal threshold voltage, some transistors would likely fail to conduct current as designed. For example, for a typical 65 nm SRAM circuit, halving the supply voltage from the nominal level to 0.5V typically increases the failure rate by about 5 orders of magnitude (see Figure 1 from [16]). Since the relationship between voltage and the log of the failure is approximately linear, the error as a function of supply voltage v is approximately of the form of $\epsilon(v) = c^{-v}$, for some positive constant c . Using the fact that the energy is proportional to the square of the supply voltage [10], the energy used by a 65nm SRAM with failure rate ϵ is proportional to $\Theta(\log^2(1/\epsilon))$.

The traditional design approach to achieving fault tolerance is to set the supply voltage to be sufficiently high so that with sufficiently high probability no transistor fails. Near-

Threshold Computing simply means that the supply voltages are designed to be closer to the threshold voltage, which can potentially offer significant improvements in energy efficiency, provided another, more energy-efficient, solution for the fault-tolerance issue can be found.

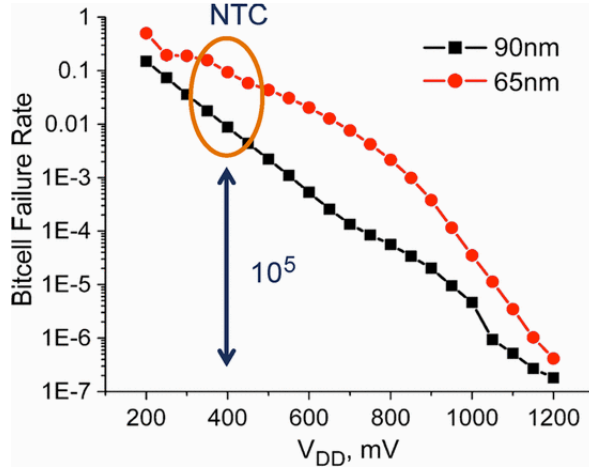


Figure 1: Semi-log plot of voltage-to-failure for an SRAM cell from [16].

One strategy to achieve fault-tolerance is to design fault-tolerant circuits, namely circuits that correctly compute the desired output if the number of failures is not significantly higher than the expected number of failures. The study of fault-tolerant circuits is not new. Starting with the seminal paper by von Neumann [30], several papers [14, 15, 17, 18, 25–27] have considered the question of how many faulty gates, each (independently) having a (small) fixed probability of failure, are required to mimic the computation of an ideal circuit with some desired probability of correctness. In general, as the probability of gate failure increases, one would expect that more gates will be required to achieve a fixed probability of failure for the circuit. As an example from [16], the circuit shown in Figure 2a is the traditional 6-transistor design for an SRAM cell, while the circuit shown in Figure 2b is a more fault-tolerant, and thus more suited for Near-Threshold Computing, 10-transistor design for an SRAM cell.

Our goal here is to initiate the theoretical study of the design of energy-efficient circuits. We assume that the design of the circuit specifies both the circuit layout as well as the supply voltages for the gates. To obtain maximum energy efficiency, the circuit design must balance the conflicting demands of minimizing the energy used per gate, and minimizing the

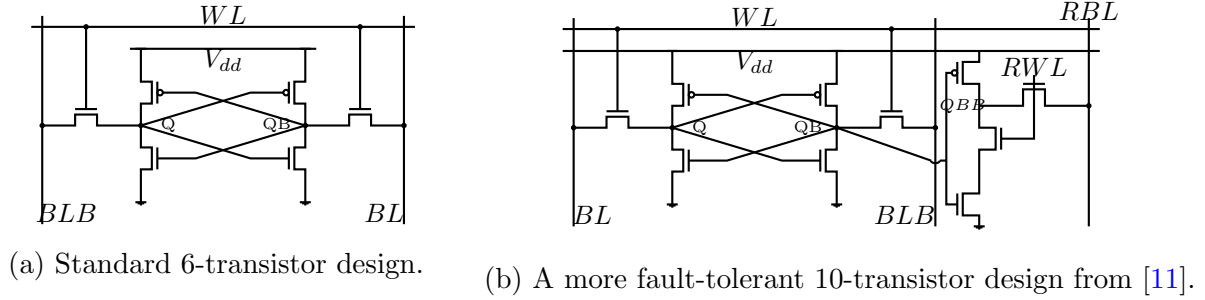


Figure 2: Two SRAM circuits with the same functionality.

number of gates in the circuit; If the energy supplied to the gates is small, then functional failures are likely, necessitating a circuit layout that is more fault-tolerant, and thus has more gates. Thus the design should find a “sweet spot” for the supply voltages that balances the competing demands of small circuit size and low per-gate energy.

Perhaps the most natural question that arises within this context is the following: Given a function, what is the minimum energy required by a circuit to compute that function? Answering this involves finding the optimal tradeoff between reduced supply voltage and increased circuit size for that function, which is closely related to the problem of finding the smallest circuit to compute that function. Due to this relationship, coupled with the fact that the ability to determine the smallest circuit to compute an arbitrary function would have vast implications throughout theoretical computer science (for example, proving that an NP-complete problem requires circuits of super-polynomial size would show $P \neq NP$), finding the optimal energy to compute an arbitrary function seems to be untouchable at this time; However, even conditional, approximate answers would provide useful insight, especially for the simpler classes of functions that circuits in computers are typically composed of. Additionally, one may be interested in optimizing the energy used by a specific circuit. In this case, the question is: Given a circuit, what is the lowest possible supply voltage such that the circuit still computes correctly (with sufficiently high probability)?

In order to begin to answer these questions, we formalize them as the following problems:

Definition 1. Minimum Energy Circuit Problem (MEC): *Given a function (or relation) F , and an error bound δ , output a circuit layout C and a setting v of the supply voltage, such that C uses minimal energy, subject to the constraint that C computes F with probability at least $1 - \delta$.*

Definition 2. Minimum Circuit Energy Problem (MCE): *Given a circuit layout C , and an error bound δ , output a setting v of the supply voltage, such that C uses minimal energy, subject to the constraint that C computes correctly (what C would compute if there were no errors) with probability at least $1 - \delta$.*

While it may not currently be practical, in principle the supply voltages need not be homogeneous over all gates of a circuit, that is, different gates could be supplied with different voltages. This naturally leads to the question of whether allowing heterogeneous supply voltages might yield lower-energy circuits than is possible if the supply voltages are required to be homogeneous. Intuitively, heterogeneous supply voltages should benefit a circuit where certain parts of the computation are more sensitive to failure than others. This naturally leads to the heterogeneous versions of the above two problems:

Definition 3. Heterogeneous Minimum Energy Circuit Problem (HMEC): *Given a function (or relation) F , and an error bound δ , output a circuit layout C and a setting v_g of the supply voltage for each gate $g \in C$, such that C uses minimal energy, subject to the constraint that C computes F with probability at least $1 - \delta$.*

Definition 4. Heterogeneous Minimum Circuit Energy Problem (HMCE): *Given a circuit layout C , and an error bound δ , output a setting v_g of the supply voltage for each gate $g \in C$, such that C uses minimal energy, subject to the constraint that C computes correctly (what C would compute if there were no errors) with probability at least $1 - \delta$.*

A variety of questions can be formulated around these problems. Perhaps most basically, for which functions can we solve MEC and HMEC (exactly or approximately)? As discussed, there are a number of functions for which solving this question seems to be beyond the reach of even state-of-the-art mathematics. However, even if it is infeasible to solve MEC and HMEC for most functions, it may be possible to say something about the minimum energy required to compute a randomly chosen function: If a function is chosen uniformly at random,

how much energy does the solution to MEC or HMEC require (with high probability)? As these questions are formulated for both MCE and HMCE, it is natural to consider how much supply voltage heterogeneity effects the solutions: For which functions is the solution to HMEC significantly less than the solution to MEC? And for a function chosen uniformly at random, is the solution to HMEC significantly less than the solution to MEC? Switching to the problems for determining the minimum energy required by a fixed circuit, we can ask for which circuits can we solve MCE and HMCE (exactly or approximately)? In this case it is clear that HMCE can have a significantly smaller solution than MCE, as a fixed circuit may contain a large “useless” component that does not affect circuit computation, and thus heterogeneous supply voltages may be employed to spend little or no energy on that part of the circuit. However, it is still interesting to consider whether, for a very natural circuit that computes some function or relation, the solution to HMCE is significantly smaller than the solution to MCE.

1.1 RELATED WORK

As far as we are aware, this is the first work to consider energy-efficient circuit design from a theoretical perspective; However, there has been a significant amount of related work in the area fault-tolerant circuit design, which we summarize here. Von Neumann [30] first introduced the model where each gate of the circuit fails with some independent, fixed probability ϵ . Von Neumann also informally argued that any function that can be computed by a faultless circuit of size s can be computed by a faulty circuit of size $O(s \log s)$. Dobrushin and Ortyukov [15] proved this result formally. Subsequently, Pippenger [25] improved this result by giving explicit, rather than probabilistic, constructions of certain aspects of the proof, resulting in an explicit construction of a fault-tolerant circuit for some fixed values of ϵ . Finally, Gács [17] proved this result in full generality for arbitrary values of ϵ . Pippenger [25] also proved that any function can be computed by a network of $O(2^n/n)$ faulty gates. When combined with Shannon’s result that almost all functions require $\Omega(2^n/n)$ faultless gates [28], this shows that almost all functions require only a constant factor increase in

circuit size when the gates are faulty.

The remaining related work is in developing lower bounds on faulty circuit size. Dobrushin and Ortyukov [14] attempted to show that any function of sensitivity m requires at least $\Omega(m \log m)$ faulty gates to compute, however their proof contains errors, which were first pointed out by Pippenger et al. [26]. Gács and Gál [18] proved the lower bound correctly, using a similar idea to [14] but proving new lemmas for the analysis. Reischuk and Schmeltz [27] independently gave another proof of the lower bound using decision trees.

The general idea of trading accuracy of a hardware circuit and computing architecture for energy savings dates back to at least [24]. The paper [16] gives an excellent survey on Near-Threshold Computing. According to [16], the three main barriers to the widespread use of Near-Threshold computing are:

1. **Performance Loss:** Circuits supplied a Near-Threshold voltage perform orders of magnitude slower than circuits supplied the nominal voltage.
2. **Increased Performance Variation:** Circuits supplied the nominal voltage experience a roughly 1.5X performance variation, while circuits supplied a Near-Threshold voltage experience up to a 20X performance variation.
3. **Increased Functional Failure:** Circuits experience an increased sensitivity to process, temperature, and voltage, resulting in an increased rate of functional failures.

While significant research has been performed to mitigate all of these barriers, the work presented in this dissertation is meant to provide the theoretical basis for solving the third problem, that of increased functional failure. From a system-design perspective, the problem of increased functional failures has been most pronounced in SRAM, and thus much of the research on this problem has been focused on making SRAM robust when supplied a Near-Threshold voltage [16]. As previously mentioned, some research has been in designing SRAM bitcells that are more fault tolerant, for example with 8 or 10 transistors, rather than the standard 6 transistor design [11, 12]. Other work has been in designing SRAM cache architectures using error-correcting codes, redundancy, and other methods to ensure cache reliability, at the cost of increased cache size or latency [1, 3, 13, 21, 22, 31, 32].

As an example of a technology that has at least the spirit of Near-Threshold Computing,

IBM’s production POWER7 servers use a technique called *Guardband* to save energy by dynamically lowering operating voltage [4].

1.2 OUR CONTRIBUTIONS

We discuss our main results in the following sections.

1.2.1 General Bounds on Circuits With Constant Error

We begin in Chapter 3 by showing general lower and upper bounds on the amount of energy required by a circuit to compute a given relation, when the reliability parameter δ is a fixed constant.

General lower bound on the energy to compute a function: We first show in Section 3.1 the following lower bound on the amount of energy required to compute any relation:

Theorem 5. *Let $\delta < 1/4$, and let C be a circuit that computes a relation h of sensitivity m with probability at least $1 - \delta$. Then C requires $\Omega\left(m \log\left(m \frac{1-2\sqrt{\delta}}{\delta}\right)\right)$ energy.*

Gács and Gál [18], and independently Reischuk and Schmeltz [27], show that any Boolean function f with sensitivity m (roughly the number of input bits which affect the output) requires a circuit of size $\Omega(m \log m)$ to be reliably computed when the gates of the circuit fail independently with a fixed positive probability. We modify the techniques in [18] to prove our lower bound on the energy required by any circuit that computes a relation with sensitivity m . The proof consists of two main parts. The first part is to consider a failure model that is equivalent in terms of the reliability of any part of the circuit, but where failures occur, and energy is consumed, on the wires of the circuit as well as at the gates. The second part considers the sensitive input bits to the circuit: If there are too few wires emanating from these input bits, then the probability that failures will cause the circuit to compute as if one of the sensitive bits were flipped is too large, and the output would be incorrect with too high a probability. An additional technical hurdle to obtaining our result

is that supply voltages may be heterogeneous, so, unlike the setting of the previous work, different gates of the same circuit may use different amounts of energy, and thus fail with different probabilities.

General upper bound on the energy to compute a function: In Section 3.2, we extend the classic upper bound on circuit size from the fault-tolerant circuit literature to obtain the following upper bound on circuit energy consumption:

Theorem 6. *Given a reliable circuit C of size s , and a fixed constant $\delta > 0$, it is possible to construct a circuit C' with homogeneous voltage supplies that uses $O(s \log(s))$ energy and that computes the same function computed by C with probability at least $1 - \delta$.*

Von Neumann [30] showed that given a Boolean function f and a circuit of size s which computes f , a circuit of size $O(s \log s)$ is sufficient for computing f correctly with high probability when the gates of the circuit fail independently with a fixed positive probability. Using techniques from [30] and from [17, 25], we show that a relation h that is computable by a circuit of size s can, with probability at least $1 - \delta$, be computed by a circuit of faulty gates using $O(s \log(s))$ energy. In our construction, the supply voltages are homogeneous. The construction works by introducing a $\Theta(\log s)$ factor of redundancy in the circuit, and each gate of the circuit is replaced by a gadget. The input to each gadget is $\Theta(\log s)$ wires per original gate input, most of which, with high probability, carry the same input bit as if the computation were being performed in the original, faultless circuit. The gadget contains $\Theta(\log s)$ copies of the original gate, as well as a component of size $\Theta(\log s)$ that ensures that the fraction of incorrect wires exiting the gadget is sufficiently low with high probability.

1.2.2 Introduction to Heterogeneity

In Chapter 4, we consider simple cases in which allowing heterogeneous supply voltages both does and does not yield asymptotic decreases in energy.

Settings where heterogeneous supply voltages are not beneficial: In Section 4.1 we observe that, when δ is restricted to a fixed constant, there are relations, namely the parity function, for which allowing heterogeneous supply voltages will not allow one to achieve a circuit design that uses asymptotically less energy than is achievable by a circuit design with

homogeneous supply voltages:

Theorem 7. *Let $\delta < 1/4$ be a fixed constant. The energy used by any circuit that computes the parity function with probability $1 - \delta$ is $\Omega(n \log(n))$, and this is achievable by a circuit with homogeneous supply voltages.*

Intuitively, the parity function has such high sensitivity that every gate in any reasonable circuit will be of equal importance, so nothing can be gained by heterogeneous supply voltages. This immediately implies that in the setting where δ is a fixed constant, heterogeneity doesn't significantly benefit some functions. That is, for these functions the optimal solution to MEC and the optimal solution to HMEC use asymptotically the same energy (up to constants), which also implies that there are circuits for these functions such that the optimal solution to MCE and the optimal solution to HMCE use asymptotically the same energy. Formally, the proof is a corollary of our lower and upper bounds from Sections 3.1 and 3.2.

A setting where heterogeneous supply voltages is beneficial: In contrast, in Section 4.2 we give a natural example where allowing heterogeneous supply voltages allows one to use asymptotically less energy than would be achievable using homogeneous supply voltages. In particular, we consider a natural super-majority relation called LSR, which outputs the majority of the input bits if this majority is sufficiently large, and the most natural circuit that computes this relation, a balanced tree of majority gates, and obtain the following theorem:

Theorem 8. *Let $E_1(\delta)$ be the optimal energy consumption of the majority tree on n leaves with homogeneous supply voltages that computes LSR with probability $1 - \delta$, and $E_2(\delta)$ be the optimal energy consumption of the same if supply voltages may be heterogeneous. Then, for $\delta' = \frac{2}{\log_3 n}$, it holds that $\frac{E_1(\delta')}{E_2(\delta')} = \omega(1)$.*

We show that for homogeneous supply voltages the energy required by this circuit to compute LSR is $\Omega(n \log^2(\delta))$, where n is the number of input bits. We then show that if supply voltages can be heterogeneous, this circuit can compute LSR using energy $O(n + 3^{1/\delta} \log^2(\delta/10))$, which is asymptotically less than $n \log^2(\delta)$ if $\delta \rightarrow 0$ as $n \rightarrow \infty$. This implies that there are quite simple relations and circuits for which the optimal solution to

HMCE uses asymptotically less energy than the optimal solution to MCE. The heterogeneous voltage setting is quite intuitive: Since we only care about a super-majority, the gates far from the output gate can sustain a small but constant fraction of failures without affecting the output, while the gates closer to the output, of which there are a comparatively small number, affect the output much more dramatically, so we use a large amount of energy to ensure they do not fail. For the proof, we lower bound the energy used by any homogeneous setting by observing that in any homogeneous setting, gates, and in particular the output gate, cannot fail with probability greater than δ . We give an explicit setting of voltages for the heterogeneous upper bound, and use recurrence relations to bound the probabilities of “bad” failure profiles, to show that, with that heterogeneous setting, the circuit outputs correctly with probability at least $1 - \delta$.

1.2.3 Hardness Results

Chapter 5 discusses complexity theoretic barriers to energy-efficient circuit design. Ideally, a circuit designer would like to solve MEC, but, for many functions, the ability to even approximately bound optimal circuit sizes is essentially at least as hard as the P vs. NP question,¹ and is untouchable with current mathematical knowledge. In order to avoid this mathematical barrier, we instead consider MCE. We show that this problem is NP-hard. Thus if $P \neq NP$ then there is no efficient method for computing the optimal supply voltage setting. The standard fallback approach for NP-hard optimization problems is to seek algorithms that are guaranteed to produce solutions with optimal/good relative error compared to the optimal solution. In our case, an algorithm A has approximation ratio c (or equivalently worst-case relative error $c - 1$) if for all inputs, the energy used by the circuit with the supply voltage setting given by A is at most c times the optimal minimum energy.

The approximation ratio of the traditional approach: We show in Section 5.1 that the approximation ratio of the traditional algorithm, which sets the supply voltages such that the probability that even a single gate fails is at most δ , is $O(\log^2 s)$, where s is the number of gates in the circuit:

¹If one could prove that your favorite NP-complete problem required super-polynomially many gates to compute, this would prove $P \neq NP$.

Theorem 9. *The traditional approach is an $O(\log^2 s)$ -approximation for MCE.*

Similar to the lower bound on the energy used by homogeneous voltage settings in Section 4.2, we can upper bound the failure probability of any homogeneous setting by δ . The result follows by noting that setting the failure probability to δ/s uses $\Theta(\log^2 s)$ more energy, and that with this failure probability, the probability that even one gate fails is at most δ .

Hardness of improving upon the traditional approach: In contrast, we show in Section 5.2 that it is NP-hard to approximate MCE to within a factor polynomially less than $O(\log^2 s)$:

Theorem 10. *It is NP-hard to $O(\log^{2-\gamma} s)$ -approximate MCE for any $\gamma > 0$.*

The circuit used in the reduction is the natural circuit for a 3SAT instance. The input bits represent the boolean settings of the variables, which are negated as appropriate and fed into OR gates, representing the clauses. The output to these OR gates is then fed into a tree of AND gates, and thus if the variable assignment satisfies the 3SAT formula, all of the inputs of the AND tree are 1's, and the circuit outputs 1; Otherwise, the circuit outputs 0. Intuitively, the AND tree is very prone to failures, and is very likely to output 0 if even a few gates of the AND tree fail, regardless of the input. Because of this, in the case that the 3SAT formula is satisfiable, the failure rate must be low enough such that every gate of the AND tree works correctly with sufficiently high probability, and so the energy must be high. On the other hand, if the 3SAT formula is not satisfiable, any input to the circuit should output 0, and thus a small number of failures in the AND tree do not significantly decrease the probability that it outputs 0. We also observe here that, using essentially the same proof as MCE, one can show HMCE is at least as hard in terms of approximation as MCE, as the ability to supply the gates of the AND tree of the 3SAT circuit different voltages yields no asymptotic benefit.

Generalization of hardness to other failure models: One might be concerned that this hardness is the result of how we specifically model functional failures in circuits, rather than due to the complexity of circuits; In order to provide evidence that the hardness does indeed come from the complexity of circuits, we prove that the same result holds in another model, the *0-default model*, that is somewhat different and, perhaps, closer to how failures

occur in real circuits: In the 0-default model failures cause wires to carry a “default” 0 bit. In fact, we even show in Section 5.3 that it is NP-Hard in this model to even determine whether a circuit computes correctly on a fixed input.

Theorem 11. *In the 0-default model, it is NP-hard to $O(\log^{2-\gamma} s)$ -approximate MCE for any $\gamma > 0$. Additionally, it is NP-Hard to determine if a circuit computes correctly on a fixed input.*

Bypassing hardness via specific families of circuits: Putting these results together, we see that there is a complexity theoretic obstacle to achieving more energy efficient circuits by using lower supply voltages than one obtains by the traditional high supply voltage approach. More precisely, if one could find a computationally efficient algorithm for setting supply voltages that has better worst-case relative error than the traditional approach, then $P=NP$. So, assuming $P \neq NP$, any proposed algorithm would either not have worst-case relative error better than the traditional approach, or would take super-polynomial time on some circuits. But of course the standard caveat applies here: as NP-hardness is a worst-case concept, this doesn’t mean that one cannot beat the energy used by the traditional approach for particular circuits of interest. As a small step in the direction of showing that for circuits of interest MCE may be approachable, we show in Section 5.4 that there is an efficient algorithm to verify whether a particular setting of the supply voltage achieves the desired error bound if the circuit is a tree.

Lemma 12. *Let C be a circuit with a tree as the underlying graph, and suppose each gate g of the circuit fails with probability ϵ_g . Then there is a polynomial time algorithm to determine if C computes correctly with probability at least $1 - \delta$.*

This result hints at the hardness of MCE coming from “cycles” in the circuit. In Section 5.5 we make the curious observation that there are circuits where the reliability of the output is not monotone in the reliability of the gates. Understanding this non-monotonicity seems to be the key to being able to solve MCE for circuits that are trees.

1.2.4 Almost All Functions Require Exponential Energy

Chapter 6 discusses the minimum amount of energy required to compute almost all, i.e., a $1 - o(1)$ fraction of, functions. In principle, for every function f , MEC has some solution. Though for many functions finding the solution to MEC may be untouchable due to the previously discussed complexity-theoretic barriers, one might hope to determine the amount of energy required by an average Boolean function. Pippenger showed that all Boolean functions can be computed by circuit layouts with $O(2^n/n)$ noisy gates [25]. Using that construction, it immediately follows that all Boolean functions can be computed by some circuit that uses $O(2^n/n)$ energy assuming δ is a fixed constant. We show in Chapter 6 that this result is tight for almost all functions, i.e:

Theorem 13. *For any $0 < \delta < 1/2$, almost all Boolean functions on n variables require circuits that use $\Omega(2^n/n)$ energy.*

To develop intuition about this result, it is necessary to consider more precisely the relationship between the voltage supplied a gate and the probability that it actually fails. Let ϵ be the function mapping voltage to error probability. In the *exact failure model*, when a gate is supplied a voltage v , it fails with probability *exactly* $\epsilon(v)$. On the other hand, in the *bounded failure model*, the gate fails with probability *at most* $\epsilon(v)$. The bounded failure model is arguably more realistic, in the sense that the circuit designer may not know exactly the probability that a gate will fail, and the failure probability may vary with time or environmental conditions. On the other hand, the exact failure model allows the circuit designer to use component failures as a source of randomness, and thus perhaps perform computations more efficiently. All of our results discussed thus far have held for both exact and bounded failure models.

Circuits that can compute many functions: The main component of the proof that almost all functions require exponential energy is to show that almost all functions require circuit layouts with exponentially many gates. In the bounded failure model, this directly follows from Shannon’s result that almost all functions require circuits of exponential size [28], since in the bounded failure model a circuit must compute correctly even if no gates fail. The exact failure model is less straightforward, as, by modifying the voltage supplied to

every gate, a single circuit layout may be able to compute a number of different functions. In fact, we have the following theorem for circuits with homogeneous supply voltages:

Theorem 14. *For any $0 < \delta < 1/2$ and $n \in \mathbb{N}$, there exists a circuit with n inputs of size $O(n)$ that computes $\Omega\left(\frac{\log n}{\log(\frac{1}{\delta} \log n)}\right)$ different functions with probability at least $1 - \delta$.*

The circuit is composed of trees of AND gates of varying size, and one can see how this circuit computes multiple functions by observing that, as the failure rate increases, a tree of AND gates will switch from computing the AND function on its input bits, to computing the 0 function.

If supply voltages are allowed to be heterogeneous, we obtain the following theorem:

Theorem 15. *For any $0 < \delta < 1/2$ and $n \in \mathbb{N}$, there exists a circuit C with n inputs of size $O(n^2)$ that computes $\Omega(3^n)$ different functions with probability at least $1 - \delta$.*

The circuit in this proof is a modification on the natural circuit for a 3CNF formula: Instead of connecting the literals in each clause directly to an input or its negation, for each literal, the output of a series of AND trees (each of which may output 1 or 0, depending on the voltage settings) specifies which input variable should be represented by that literal, and whether or not it should be negated. Thus, by modifying supply voltages, the circuit is able to compute any function on n inputs that can be represented by a 3CNF formula with a fixed number of fixed-sized clauses.

Upper bounds on the number of functions a circuit can compute: Despite the existence of circuits that can compute many functions, we are able to provide, both for homogeneous and heterogeneous supply voltages, a sufficiently small upper bound on the number of functions that a single circuit can compute:

Lemma 16. *A circuit C on n inputs with s gates can compute at most $s2^n + 1$ functions if supply voltages must be homogeneous, and at most $(8e2^n)^s$ functions if supply voltages may be heterogeneous.*

Though the proof for heterogeneous supply voltages is somewhat more complicated than the one for homogeneous supply voltages, both follow primarily from two observations: (1) The probability a fixed circuit outputs a 1 on a fixed input can be written as a polynomial in the failure rate of each gate of the circuit, and (2) A circuit with a fixed setting of

supply voltages only computes a function when, on all inputs, it outputs either 1 or 0 with probability at least $1 - \delta$. Combining these two, we observe that a circuit with a set of supply voltages only computes a function when, for each polynomial associated with the probability of outputting a 1 on an input, that polynomial is above $1 - \delta$ or below δ . With this in hand, we can apply results from calculus and geometry to obtain the upper bound on the number of functions a single circuit can compute.

Equivalence of exponential energy and exponentially many noisy gates: These results leave open the possibility that some Boolean functions that do not require circuits with exponentially many gates still require exponential energy. For example it could be the case that for some function the energy optimal circuit has sub-exponentially many gates, with many of them requiring exponential energy. We show that this is not the case:

Lemma 17. *A Boolean function f requires circuits that use exponential energy if and only if it requires circuits that contain exponentially many gates.*

The proof follows by noting that, for any circuit, setting supply voltages such that the energy-per-gate is polynomial in the circuit size is sufficient for no gate in a circuit to fail, and thus using a higher energy-per-gate can provide no additional benefit.

1.2.5 The Power of Heterogeneity to Reduce Energy

In Chapter 7, we explore the power of heterogeneous supply voltages to asymptotically save energy over any circuit using homogeneous supply voltages when computing certain functions or relations. In this chapter, we consider the case when the error parameter δ vanishes as the number of the inputs to the function increases. Previously, in Chapter 3, we focused on the case when δ is a fixed constant. This intuitively makes sense, as a circuit designer may calculate the requirements for δ , and design the circuit based on that. However, it is also reasonable to consider the case when δ must vanish as the number of inputs to the circuit increases: As the circuit size increases, circuit failures may become more expensive to recover from; Additionally, a single circuit may be a component of a larger system, and as this system grows, it will not function reliably enough if the the failure probability of the components it is made up of does not decrease.

Many functions benefit from heterogeneous supply voltages: Perhaps surprisingly given the general upper and lower bounds in Chapter 3 when δ is a fixed constant, which are tight for many natural functions (e.g., parity), we show that, if the circuit error probability must vanish as the number of inputs increases, there is a natural class of functions (that includes parity) which can be computed with asymptotically less energy if supply voltages are allowed to be heterogeneous.

Theorem 18. *For any function f with minimum circuit size s , for any constant $c > 0$, if $\delta = 1/s^c$, then every circuit with homogeneous supply voltages computing f uses $\Omega(s \log^2(s))$ energy, and there exists a circuit with heterogeneous supply voltages using $O(s \log s)$ energy.*

If we replace s by $\Theta(n)$, where n is the number of inputs to the function, then the above theorem applies to all functions with circuits of size linear in the number of inputs. Thus, for such functions, the solution to HMEC is asymptotically less than the solution to MEC by a factor of $\Omega(\log n)$ when δ is polynomial in $1/n$. In order to show this, we first provide a $\Omega(s \log^2 s)$ lower bound on the energy used by any circuit using homogeneous supply voltages, which can be obtained by noting that the voltage setting must be such that the output gate does not fail with probability more than δ . On the other hand, we provide a general circuit construction and heterogeneous voltage setting for functions in this class that uses $O(s \log s)$ energy. Intuitively, in a manner similar to the upper bound in Chapter 3, we replace each gate of a faultless circuit with a fault-tolerant gadget, and supply low voltages to these gates, with the result that the failure rate for each gate of each gadget is a constant. We are left at the end with a (relatively small) set of wires, such that, with high probability, the majority of these wires carry the correct output bit. We then use a majority circuit, with voltages set sufficiently high so that with high probability no gate fails, and thus obtain the correct output bit with high enough probability.

Limit for many functions on the energy savings via heterogeneous supply voltages: We then show that for functions that have circuits of linear size and obtain a $\Omega(\log n)$ energy savings via heterogeneous supply voltages, this energy savings is tight, i.e., the solution to HMEC is only a factor of $\Theta(\log n)$ less than the solution to MEC, as long as most input bits are non-degenerate. An input bit to a function is *non-degenerate* if, roughly,

there is some input to the function where the value of that bit matters.

Theorem 19. *Let f be a function with b non-degenerate input bits. Then, for any $0 < \delta < 1/2$, any circuit C that computes f with error at most δ requires $\Omega(b \log 1/\delta)$ energy.*

We show that for any function, each input bit that is non-degenerate essentially must use $\Theta(\log n)$ energy by itself, or else the output cannot possibly be correct with high enough probability. If the number of non-degenerate bits of a function is a constant fraction of n , then even circuits with heterogeneous supply voltages computing that function must use $\Omega(n \log n)$ energy when $\delta = 1/n$.

Relations gaining greater energy savings via heterogeneous supply voltages:

In principle, a circuit may compute a relation, rather than a function, as, for example, the proper functioning of the system may guarantee that the circuit does not receive certain inputs. This raises the question of whether heterogeneous supply voltages can provide a greater energy savings over homogeneous supply voltages when computing relations rather than functions, i.e., whether a circuit with heterogeneous supply voltages that computes a relation can save $\omega(\log n)$ energy over any circuit with homogeneous supply voltages computing that relation. We answer this question in the positive.

Theorem 20. *Suppose $\delta = 1/n^c$ for some constant $c > 0$. Then there is a relation that can be computed by a heterogeneous circuit using $O(n)$ energy, but for homogeneous circuits requires $\Omega(n \log^2 n)$ energy.*

The relation that obtains the $\Theta(\log^2 n)$ energy savings by using heterogeneous supply voltages is a natural supermajority relation. This $\Theta(\log^2 n)$ energy savings is the best savings possible for computing any relation that requires a faultless circuit of size $O(n)$ that must access $\Theta(n)$ of the input bits. Our lower bound on the energy used by any circuit with a homogeneous voltage setting is similar to our previous such lower bounds. The upper bound using a heterogeneous voltage setting modifies a standard majority circuit, and benefits from the fact that, in such a circuit, failures occurring closer to the inputs have only a small effect on the output of the circuit. As we traverse down the circuit, failures become more problematic, so we increase the redundancy linearly in order to make failures more rare, but since the number of gates decreases exponentially, adding this redundancy only increases the

circuit size by a constant.

2.0 MODEL, DEFINITIONS, AND NOTATION

We now make formal definitions. A *Boolean relation* h is a map from $\{0, 1\}^n$ to $\{0, 1\}$, where each input is mapped to 0, 1, or both 0 and 1. If $x \in \{0, 1\}^n$ is mapped to both 0 and 1, this can be thought of as “don’t care” (for example because the input x should not occur in a correctly functioning system). A *Boolean function* f is a Boolean relation where each input is uniquely mapped to either 0 or 1. For any input $x \in \{0, 1\}^n$, denote by x^ℓ the input that has the same bits as x , except for the ℓ -th bit, which is flipped. A Boolean relation h is *sensitive* on the ℓ -th bit of x if neither $h(x)$ nor $h(x^\ell)$ is mapped to both 0 and 1, and $h(x) \neq h(x^\ell)$. The *sensitivity of h on x* is the number of bits of x that h is sensitive on. The *sensitivity of h* is the maximum over all x of the sensitivity of h on x .

A *gate* is a function $g : \{0, 1\}^{n_g} \rightarrow \{0, 1\}$, where n_g is the number of inputs (i.e., the *fan-in*) of the gate. We assume that the maximum fan-in is at most a constant. A *Boolean circuit* C with n inputs is a directed acyclic graph in which there are n nodes with no incoming edges that each output one of the input bits, and every other node is a gate. The *size* of a circuit, denoted by s , is the number of gates it contains. For any $I \in \{0, 1\}^n$, we denote by $C(I)$ the output of the Boolean function computed by Boolean circuit layout C .

In this paper we consider circuits (C, \bar{v}) that consist of both a traditional circuit layout C as well as a vector of supply voltages \bar{v} , one for each gate of C . Every gate g is supplied with a voltage v_g . We say that the supply voltages (and, as shorthand, circuit) are *homogeneous* when every gate of the circuit is supplied with the same voltage, and *heterogeneous* otherwise. We say that a gate *fails* when it produces an incorrect output, that is, when given an input x it produces an output other than $g(x)$.

The benefits of allowing the circuit designer to control the voltage depends both on the rate at which failures decrease as voltages increase as well as whether or not failure rates are

previously known. If the circuit designer knows exactly the probability that a component will fail when supplied a specific voltage, then this may be used as a source of randomness, which could in theory allow for more efficient computation. Because of this, we consider two different failure models. In the *exact failure model*, each non-input gate g fails independently with probability *exactly* $\epsilon(v_g)$. In contrast, in the *bounded failure model*, every non-input gate g fails with probability *at most* $\epsilon(v_g)$ (we specify the failure model only for results that do not hold for both). While the bounded failure model perhaps more closely models reality, in the sense that the actual error rate of a circuit component may be unknown or may vary with time or environment, the exact failure model gives the circuit designer more power. In both models we assume $\epsilon : \mathbb{R}^+ \rightarrow (0, 1/2)$ is a decreasing function. The voltage supplied to a gate determines both its energy usage and its failure probability, thus we define $\epsilon_g := \epsilon(v_g)$ and drop all future formal reference to supply voltages. Finally we assume there is a decreasing, nonnegative failure-to-energy function $E(\epsilon)$ that maps the failure probability ϵ to the energy used by a gate. The energy required by a circuit C is simply the aggregate energy used by the gates, $\sum_{g \in C} E(\epsilon_g)$ in our notation. Throughout the majority of this dissertation we assume $E(\epsilon) = \Theta(\log^2(1/\epsilon))$; Throughout, in the appropriate locations, we discuss how to generalize our results to other failure-to-energy functions.

A gate that never fails is said to be *reliable* or *faultless*. Given a value $\delta \in (0, 1/2)$ (δ may not be constant), a circuit $(C, \bar{\epsilon})$ that computes a Boolean relation h is said to be $(1 - \delta)$ -*reliable* if for every input I on which $h(I)$ is not both 0 and 1, $C(I)$ equals $h(I)$ with probability at least $1 - \delta$. We say that C can compute ℓ different functions $(1 - \delta)$ -reliably if there exists $\bar{\epsilon}_1, \bar{\epsilon}_2, \dots, \bar{\epsilon}_\ell \in (0, 1/2)^{|C|}$ and different functions f_1, f_2, \dots, f_ℓ such that $(C, \bar{\epsilon}_i)$ is $(1 - \delta)$ -reliable for function f_i . We say that a circuit is *reliable* or *faultless* if it is 1-reliable (for example, because all its gates are reliable). We say that the circuit is (ϵ, δ) -reliable if it is $(1 - \delta)$ -reliable when gates fail with probability exactly ϵ .

3.0 GENERAL ENERGY UPPER AND LOWER BOUNDS

In this chapter, we prove a general lower bound, in terms of sensitivity, and a general upper bound, in terms of circuit size, on the amount of energy required to compute a function.

3.1 A GENERAL ENERGY LOWER BOUND

Our main goal in this section is to prove Theorem 21, which roughly states that $\Omega(m \log m)$ energy is necessary to compute a relation with sensitivity m .

Theorem 21. *Let $\delta < 1/4$, and let C be a circuit that $(1 - \delta)$ -reliably computes a relation h of sensitivity m . If each gate g of C fails independently with probability ϵ_g , and incurs an energy consumption of $E(\epsilon_g)$, with E being a proper failure-to-energy function, then C requires*

$$\Omega\left(m \log \left(m \frac{1 - 2\sqrt{\delta}}{\delta}\right)\right)$$

energy in order to $(1 - \delta)$ -reliably compute h .

The outline of this section is as follows. First, we define proper failure-to-energy functions (Definition 22), and discuss why proper functions are natural. Then, similarly to [18], we show how to translate our problem to an equivalent problem where the failures occur not only on gates, but on wires as well. This is formalized in Statement 23, which is implied by the proof of Lemma 3.1 in [14], and is also used in [18]. Lemma 26 then gives a lower bound on the energy necessary for $(1 - \delta)$ -reliable circuits within this new model with wire failures. The proof is based on the proof of Theorem 3.1 in [18], and uses a series of inequalities that relate the probability of an input being incorrectly transmitted to the probability of the

circuit being incorrect. Using this, we can write the problem as a single-variable optimization problem and use standard techniques to give the desired lower bound. Finally, to prove Theorem 21 we show that given a $(1 - \delta)$ -reliable circuit C in our original model without wire failures, we can create a $(1 - \delta)$ -reliable circuit C' in the new model with wire failures, where the energy consumptions of C and C' differ only by a constant.

Definition 22. *A failure-to-energy function E is called proper when it satisfies the four following properties:*

1. E is nonincreasing,
2. $\lim_{\epsilon \rightarrow 0^+} E(\epsilon)/(\log 1/\epsilon) > 0$,
3. $\lim_{\epsilon \rightarrow 1/2^-} E(\epsilon) > 0$,
4. $E(\epsilon_1) + E(\epsilon_2) \geq 2E(\sqrt{\epsilon_1\epsilon_2})$ for all $\epsilon_1, \epsilon_2 \in (0, 1/2)$.

The first and third restrictions are natural, since they just require that the energy used decreases, but never becomes zero, as the probability of failure of a gate increases. The second property states that the energy must increase “quickly enough” as the probability of a gate’s failure tends to 0, which is necessary in order to have any energy saving over gates that never (or almost never) fail. The last property provides a convexity constraint on the function P . We point out that failure-to-energy functions typically observed in real gates fall within this class of proper failure-to-energy functions [16, 20].

Statement 23 ([14]). *Let g be a gate with fan-in n_g , in a circuit C where both gates and wires may fail. Furthermore, let $\epsilon \in (0, 1/2)$, $\zeta_g \in [0, \epsilon/n_g]$ and let $g(t)$ be the output of gate g assuming that its input-wires receive input t , and both g and g ’s input-wires are reliable. Then there exists a unique value $\eta_g(y, \zeta_g) \in [0, 1]$ such that if*

- *the input wires of g fail independently with probability ζ_g , and*
- *gate g fails with probability $\eta_g(y, \zeta_g)$ when the gate receives input y ,*

then the probability that g does not output $g(t)$ is equal to ϵ .

Note that in Statement 23, since we can now have failures on wires, the input y received by a gate g may be different than the input t received by the corresponding wires.

We need the following definition and technical lemma.

Definition 24. Given $x_{1,1}, x_{1,2}, \dots, x_{1,n} \in \mathbb{R}$, we recursively define a sequence of numbers as follows. Let $m_j^u = \arg \max_i x_{j,i}$ and $m_j^l = \arg \min_i x_{j,i}$. Then, for all $i \neq \{m_j^u, m_j^l\}$, let $x_{(j+1),i} = x_{j,i}$, and let $x_{(j+1),m_j^u} = x_{(j+1),m_j^l} = \sqrt{x_{j,m_j^u} x_{j,m_j^l}}$.

Lemma 25. Let a_1, a_2, \dots be a sequence of numbers such that $a_j = x_{j,m_j^u} - x_{j,m_j^l}$, with the terms x_{j,m_j^u} and x_{j,m_j^l} as defined above. Then,

$$\lim_{j \rightarrow \infty} a_j = 0.$$

Proof. First note that within the $n+1$ -th recursive step of the above construction there must be some index i^* that has been chosen as m_j^l twice. Let k_1 and k_2 denote the recursive step during which i^* is chosen for the first and second time, respectively. More formally, consider $S_l = \bigcup_{j \leq l} \{m_j^l\}$. Then, k_2 is the minimum index $j \leq n+1$ such that $\{m_j^l\} \cap S_{j-1} \neq \emptyset$, and k_1 is the index $j < k_2$ such that $m_{k_2}^l \cap S_j \neq \emptyset$.

For notational convenience, we denote $x_{k_1, m_{k_1}^u}$ with x_h and $x_{k_1, m_{k_1}^l}$ with x_l . Note that the sequence formed by x_{j, m_j^u} is monotonically decreasing in j , and similarly x_{j, m_j^l} is monotonically increasing in j . Then,

$$a_{k_2} = x_{k_2, m_{k_2}^u} - x_{k_2, m_{k_2}^l} = x_{k_2, m_{k_2}^u} - \sqrt{x_h x_l} \leq x_h - \sqrt{x_h x_l}.$$

Therefore,

$$\frac{a_{k_2}}{a_{k_1}} \leq \frac{x_h - \sqrt{x_h x_l}}{x_h - x_l} = \frac{\sqrt{x_h} (\sqrt{x_h} - \sqrt{x_l})}{x_h - x_l} = \frac{\sqrt{x_h}}{\sqrt{x_h} + \sqrt{x_l}}.$$

Rewriting this, we obtain $a_{k_2} \leq a_{k_1} / (1 + \sqrt{x_l/x_h})$. Then, by observing that the sequence of a_i 's is monotonically decreasing, because the sequence formed by x_{j, m_j^u} is monotonically decreasing and x_{j, m_j^l} is monotonically increasing in j , we conclude that

$$a_{k_2} \leq \frac{a_1}{\left(1 + \sqrt{x_{m_1^l}/x_{m_1^u}}\right)}.$$

It follows that, for any positive integer x , $a_{xn+1} \leq a_1 / (1 + \sqrt{x_{m_1^l}/x_{m_1^u}})^x$, and therefore $\lim_{j \rightarrow \infty} a_j = 0$. □

Lemma 26. *Let E be a proper failure-to-energy function, and let C be a circuit that $(1 - \delta)$ -reliably computes a relation h of sensitivity m . If (i) each gate g of C fails independently with probability $\eta_g(y, \zeta_g)$ when receiving input y , (ii) g incurs an energy consumption of zero, and (iii) each wire i entering g fails independently with probability $\zeta_g \in (0, 1/4)$ and incurs an energy usage of $f(\zeta_g)$, then C requires*

$$\Omega\left(m \log\left(m \frac{1 - 2\sqrt{\delta}}{\delta}\right)\right)$$

energy in order to $(1 - \delta)$ -reliably compute h .

Proof. We start by rephrasing our problem after borrowing a constraint on the number of wires and some notation from [18]. Specifically, let z be an input such that h has maximum sensitivity on z . Let $S \subset \{1, 2, \dots, n\}$ be the set of indexes so that $\ell \in S$ if and only if h is sensitive to the ℓ -th bit on input z . Then $|S| = m$, where m is the sensitivity of h . For each $\ell \in S$ denote by B_ℓ the set of all wires originating from the ℓ -th input of the circuit. Let $w_\ell = |B_\ell|$. For any set $\beta \subset B_\ell$, let $H(\beta)$ be the event that the wires belonging to β fail and the other wires of B_ℓ are correct. Denote by β_ℓ the subset of B_ℓ where

$$\max_{\beta \subset B_\ell} \Pr[C(z^\ell) = h(z^\ell) \text{ s.t. } H(\beta)]$$

is obtained, where $C(z^\ell)$ is a random variable for the output of the circuit given input z^ℓ . Finally, let $H_\ell = H(B_\ell \setminus \beta_\ell)$. Note that since wires can now fail with different probabilities, we have that,

$$\Pr[H_\ell] = \prod_{i \in \beta_\ell} (1 - \zeta_i) \prod_{i \notin \beta_\ell} \zeta_i \geq \prod_{i \in B_\ell} \zeta_i.$$

It follows from Inequalities (5) and (6) of [18] that

$$\frac{\delta}{1 - 2\sqrt{\delta}} \geq \sum_{\ell \in S} \prod_{i \in B_\ell} \zeta_i$$

and as in [18], using the inequality of arithmetic and geometric means, we have

$$\frac{\delta}{1 - 2\sqrt{\delta}} \geq m \left(\prod_{\ell \in S, i \in B_\ell} \zeta_i \right)^{1/m}.$$

Rewriting this to isolate the product term, we have

$$\prod_{\ell \in S, i \in B_\ell} \zeta_i \leq \left(\frac{\delta}{m(1 - 2\sqrt{\delta})} \right)^m.$$

Therefore, minimizing the energy consumption, is equivalent to the following optimization problem,

$$\begin{aligned} & \text{minimize} && \sum_{\ell \in S, i \in B_\ell} E(\zeta_i) \\ & \text{subject to} && \prod_{\ell \in S, i \in B_\ell} \zeta_i \leq \left(\frac{\delta}{m(1 - 2\sqrt{\delta})} \right)^m. \end{aligned}$$

Now, take some feasible solution ζ^* to the above optimization problem. Let ζ_1^* and ζ_2^* denote the minimum and maximum ζ_i^* respectively, and M denote the total number of wires, i.e., $M = \sum_{\ell \in S} w_\ell$. Note that since we assume that $E(p_1) + E(p_2) \geq 2E(\sqrt{p_1 p_2})$ for all $p_1, p_2 \in (0, 1/2)$, we can set $\zeta_1^* = \zeta_2^* = \sqrt{\zeta_1^* \zeta_2^*}$, without increasing the value of the objective, and further the constraint remains feasible. By Lemma 25 this process, if repeated, will converge to a solution where all ζ_i are equal. Therefore, we can rewrite the optimization problem as

$$\begin{aligned} & \text{minimize} && ME(x) \\ & \text{subject to} && x^M \leq \left(\frac{\delta}{m(1 - 2\sqrt{\delta})} \right)^m. \end{aligned}$$

Isolating M in the constraint above, the problem is equivalent to that of minimizing

$$\left(\frac{m}{\log 1/x} \log \left(m \frac{1 - 2\sqrt{\delta}}{\delta} \right) \right) E(x).$$

Since the function satisfies properties 1, 2, and 3 of Definition 22, the above expression will be minimized either at some constant $x \in (0, 1/4)$, in which case $E(x)/\log(1/x) > 0$, or in the limit as x approaches 0, in which case

$$\lim_{x \rightarrow 0^+} E(x)/\log(1/x) > 0,$$

or in the limit as x approaches $1/4$, in which case

$$E(x)/\log(1/x) > 0.$$

The lemma follows. □

We are now ready to prove Theorem 21.

of Theorem 21. We start by constructing a new circuit C' for computing h , which is identical to C except that both wires and gates may fail, wires of C' incur some non-zero energy consumption (as a function of their probability of failure), and the gates in C' do not consume energy. First we argue that this can be done such that C' is $(1 - \delta)$ -reliable. Observe that if for each wire i entering gate g we set its probability of failure to $\zeta_g = \epsilon_g/n_g$, we can apply Statement 23 and set the failure probability on gate g when receiving input y to $\eta_g(y, \zeta_g)$. The result is that when the input wires of gate g in C' receive input t , the probability that g does not output $g(t)$ is ϵ_g (the same as the probability of failure of g in the original circuit C). Thus by setting these failure probabilities for each gate and wire in C' we have that, for any input x , C and C' output $h(x)$ with the same probability, and so C' is $(1 - \delta)$ -reliable.

Now we set the energy consumption of the wires such that the energy of C' is at most the energy of C . First observe that if for each gate g we set the failure-to-energy function of the wires that are inputs to g to be $\tilde{E}_g(\zeta) = E(n_g \cdot \zeta)/n_g$, then since $\zeta_g = \epsilon_g/n_g$, the total energy of the wires entering g would be $n_g \tilde{E}_g(\zeta_g) = E(\epsilon_g)$ and the energy of C and C' would be equal. However, to apply Lemma 26, all wires must have the same failure-to-energy function. Therefore, let n_g^* be the maximum fan-in of any gate of C , i.e., $n_g^* = \max_{g \in C} n_g$. We set the failure-to-energy function of all wires in C' to be

$$\tilde{E}(\zeta) = \begin{cases} E(n_g^* \cdot \zeta)/n_g^* & \text{if } \zeta < \frac{1}{2n_g^*}, \\ \lim_{\epsilon \rightarrow 1/2^-} E(\epsilon)/n_g^* & \text{if } \zeta \geq \frac{1}{2n_g^*}. \end{cases}$$

First observe that $\tilde{E}_g(\zeta) \geq \tilde{E}(\zeta)$ for all $\zeta \in (0, 1/2)$ since E is nonincreasing so $E(n_g \zeta) \geq E(n_g^* \zeta)$. This implies that the energy of C' is at most the energy of C .

In order to apply Lemma 26, we need to verify that \tilde{E} is a proper failure-to-energy function. The first property follows directly from the definition of \tilde{E} . For the second property, observe that

$$\begin{aligned} \lim_{\zeta \rightarrow 0^+} \frac{\tilde{E}(\zeta)}{\log\left(\frac{1}{\zeta}\right)} &= \frac{1}{n_g^*} \lim_{\zeta \rightarrow 0^+} \frac{E(n_g^* \zeta)}{\log\left(\frac{1}{n_g^* \zeta}\right)} \cdot \lim_{\zeta \rightarrow 0^+} \frac{\log\left(\frac{1}{n_g^* \zeta}\right)}{\log\left(\frac{1}{\zeta}\right)} \\ &= \lim_{\zeta \rightarrow 0^+} \frac{E(n_g^* \zeta)}{\log\left(\frac{1}{n_g^* \zeta}\right)} > 0. \end{aligned}$$

The third property follows from the fact that

$$\lim_{\zeta \rightarrow 1/2^-} \tilde{E}(\zeta) = \lim_{\epsilon \rightarrow 1/2^-} E(\epsilon)/n_g^* > 0,$$

where we exploited the definition of \tilde{E} and the fact that, by hypothesis, E is a proper failure-to-energy function. For the fourth property, let $\zeta_1, \zeta_2 \in (0, 1/2)$, and, w.l.o.g., $\zeta_1 < \zeta_2$. There are four cases, depending on the relationship between ζ_1, ζ_2 , and n_g^* . When $\zeta_1 < \zeta_2 < 1/2n_g^*$, by applying the definition of \tilde{E} and since E by hypothesis is a proper failure-to-energy function, we have

$$\begin{aligned} \tilde{E}(\zeta_1) + \tilde{E}(\zeta_2) &= \frac{E(n_g^* \zeta_1)}{n_g^*} + \frac{E(n_g^* \zeta_2)}{n_g^*} \\ &\geq 2 \frac{E(\sqrt{n_g^* \zeta_1 n_g^* \zeta_2})}{n_g^*} \\ &= 2\tilde{E}(\sqrt{\zeta_1 \zeta_2}). \end{aligned}$$

When $\zeta_1 < \zeta_2 = 1/2n_g^*$, by the previous case we have that

$$\lim_{\zeta_2 \rightarrow (1/(2n_g^*))^-} \left(\tilde{E}(\zeta_1) + \tilde{E}(\zeta_2) - 2\tilde{E}(\sqrt{\zeta_1 \zeta_2}) \right) \geq 0,$$

and so in this case the property holds. When $\zeta_1 < 1/2n_g^* < \zeta_2$, we have that

$$\begin{aligned} \tilde{E}(\zeta_1) + \tilde{E}(\zeta_2) &= \tilde{E}(\zeta_1) + \tilde{E}\left(\frac{1}{2n_g^*}\right) \\ &\geq 2\tilde{E}\left(\sqrt{\frac{\zeta_1}{2n_g^*}}\right) \\ &\geq 2\tilde{E}(\sqrt{\zeta_1 \zeta_2}), \end{aligned}$$

where the first equality holds by definition of \tilde{E} , the first inequality follows by the preceding case, and the second inequality holds since \tilde{E} is nonincreasing and since, in this case, $\sqrt{\zeta_1/2n_g^*} \leq \sqrt{\zeta_1 \zeta_2}$. Finally, when $1/2n_g^* \leq \zeta_1 < \zeta_2$, $\sqrt{\zeta_1 \zeta_2} > \zeta_1$ and thus, by definition, $\tilde{E}(\zeta_1) = \tilde{E}(\zeta_2) = \tilde{E}(\sqrt{\zeta_1 \zeta_2})$. We conclude that \tilde{E} is a proper failure-to-energy function. The theorem then directly follows by applying Lemma 26 to C' and \tilde{E} . \square

3.2 A GENERAL ENERGY UPPER BOUND

Our main goal in this section is to prove Theorem 27, which roughly states that $O(s \log s)$ energy is sufficient to simulate a circuit of size s .

Theorem 27. *Given a reliable circuit C of size s , a non-trivial failure-to-energy function, and a fixed constant $\delta > 0$, it is possible to construct a circuit C' with homogeneous voltage supplies that uses $O(s \log(s/\delta))$ units of energy and that $(1 - \delta)$ -reliably computes the same function computed by C .*

To prove Theorem 27 we use an upper bound on the number of gates for fault-tolerant circuits originally stated by Pippenger [25] and later proved in full generality by Gács [17]. This upper bound is stated in Theorem 28. The energy upper bound follows by choosing the voltage supply that minimizes the product of the total number of gates in the circuit constructed in Theorem 28, and the energy used by each gate. More specifically, we want to set the gate failure probability ϵ so as to minimize $E(\epsilon)/(\log(1/\epsilon) - r_0)$ for some constant r_0 . As long as E is *non-trivial*, i.e., if for any $p^* \in (0, 1/2)$ it holds that $E(p^*) < +\infty$, one can find an ϵ such that $E(\epsilon) = O(1)$. This setting of ϵ then implies the upper bound of $O(s \log(s/\delta))$ on the energy used by this construction using homogeneous supply voltages.

Theorem 28 ([17]). *There are constants $R_0, \epsilon_0, r_0 > 0$ such that for all $\epsilon < \epsilon_0$ and $\delta \geq 3\epsilon$, for every reliable circuit C of size s there is a circuit of size $R_0 \frac{s \log(s/\delta)}{\log(1/\epsilon^*) - r_0}$ that computes the same result as C with probability at least $1 - \delta$ if gates fail independently with probability at most ϵ , where $\epsilon^* = \max\{\epsilon, \delta/s\}$.*

Proof. The main idea of this construction is to replace each gate of the reliable circuit with a gadget in the fault-tolerant circuit. A constant k is chosen as the level of redundancy for the circuit, meaning that each gadget has k outputs and each input to a gate in the reliable circuit is replaced by k inputs to a gadget. Another constant $\theta \in (0, 1)$ is chosen such that, with high probability, θk wires exiting each gadget carry the same value as the corresponding gate in the reliable circuit. Each gadget contains k copies of the corresponding gate from the reliable circuit, as well as an additional circuit that ensures that at least a θ fraction of the wires exiting the gadget are correct. \square

4.0 INTRODUCTION TO SUPPLY VOLTAGE HETEROGENEITY

In this chapter, we show situations where allowing supply voltages to be heterogeneous rather than homogeneous both does and does not allow for asymptotic decreases in energy consumption. In Section 4.1, we show that, when δ is a fixed constant, there are functions which do not use asymptotically less energy when supply voltages may be heterogeneous. In contrast, in Section 4.2, we show that for the natural circuit computing a supermajority relation, allowing supply voltages to be heterogeneous allows the relation to be computed with asymptotically less energy than if the supply voltages must be homogeneous.

4.1 SUPPLY VOLTAGE HETEROGENEITY MAY NOT HELP

In this section we observe in Theorem 29 that there are relations, namely the parity function, where heterogeneous supply voltages do not allow for an asymptotic reduction in energy.

Theorem 29. *Let $\delta < 1/4$ be a fixed constant. The energy used by any circuit to $(1 - \delta)$ -reliably compute the parity function is $\Omega(n \log(n/\delta))$, and this is achievable by a circuit with homogeneous supply voltages.*

Proof. The parity function can be reliably computed by a perfect binary tree of $2n - 1$ XOR reliable gates. Thus, by Theorem 27 there exists a $(1 - \delta)$ -reliable circuit for the parity function that uses homogeneous voltage supplies and that incurs $O(n \log(n/\delta))$ energy consumption. Since the sensitivity of the parity function is n , by Theorem 21 this is the best possible to within a constant factor. \square

4.2 SUPPLY VOLTAGE HETEROGENEITY CAN HELP

The goal of this section is to prove Theorem 33, which roughly states that heterogeneous supply voltages allow the natural majority circuit to compute a super-majority with asymptotically less energy than is possible with homogeneous supply voltages.

We start by defining the circuit and the logarithmic supermajority relation (LSR). Lemma 37 shows that $\Omega(n \cdot E(\delta))$ energy is required to $(1 - \delta)$ -reliably compute the LSR with homogeneous voltage supplies. The intuition behind the proof is that the output gate of any $(1 - \delta)$ -reliable circuit cannot have a probability of failure greater than δ and, since the voltage supplies are homogeneous, neither can any other gate. Then in Lemma 38 we show that there is a heterogeneous setting of the supply voltages so that this circuit $(1 - \delta)$ -reliably computes LSR with energy $O(n + 3^{1/\delta} E(\delta/10))$. Intuitively, we split the circuit into an “upper” part consisting of gates close to the output gate, and a “lower” part consisting of gates close to the input gates: Each gate in the lower part has a constant probability of failure and thus a small energy consumption which results in non-constant savings compared to the optimal homogeneous setting. With the help of technical Lemmas 34 and 35, we are able to show that although there exist gates in the lower part of the circuit that fail with a probability higher than δ , no such gate fails with a probability $o(1)$. This preserves enough information for the upper part of the circuit to still $(1 - \delta)$ -reliably compute LSR. In other words, in the upper part of the circuit we use a much smaller probability of failure for each gate in order to ensure that the circuit will “autocorrect” itself and output the correct result (see Lemma 36).

Definition 30. *The Logarithmic Supermajority Relation (LSR) is the following Boolean relation:*

$$LSR(x) = \begin{cases} 0 & \text{if the number of 0's in } x \text{ is at least } n - \frac{1}{2} \log_3 n, \\ 1 & \text{if the number of 1's in } x \text{ is at least } n - \frac{1}{2} \log_3 n, \text{ and} \\ 0 \text{ and } 1 & \text{otherwise,} \end{cases}$$

where x is the input and $|x| = n$.

This relation outputs 1 when the input contains at least $n - (1/2) \log_3 n$ ones, 0 when the input contains at least $n - (1/2) \log_3 n$ zeros, and otherwise we “don’t care”.

Definition 31. A majority tree is a Boolean circuit where the gates form a perfect ternary tree in which the leaves represent the inputs, and each internal gate, called majority gate, outputs the majority of its three children.

Definition 32. A failure-to-energy function E is called easy-going when the following hold:

- $\lim_{x \rightarrow 0} E(x) = +\infty$
- There exists a constant $c > 0$ such that $\frac{E(x/10)}{E(x)} \leq c$ for all $x \in (0, 1/2)$.

Note that this class of failure-to-energy functions contains many natural functions. For example $E(x) = 1/x^\alpha$ and $E(x) = (\log 1/x)^\alpha$ are both easy-going.

Theorem 33. Let E be an easy-going failure-to-energy function, and let $c \in (0, 1)$ be a constant. Furthermore, let $E_1(\delta)$ be the optimal energy consumption of the $(1 - \delta)$ -reliable majority tree on n leaves where all the gates must have the same failure probability, and $E_2(\delta)$ be the optimal energy consumption of the same $(1 - \delta)$ -reliable majority tree when each gate can have an arbitrary failure probability. Then, for $\delta' = \frac{1}{(1-c) \log_3 n}$, there holds

$$\frac{E_1(\delta')}{E_2(\delta')} = \omega(1).$$

Let p_i be the probability that a gate of height i in a majority tree outputs 1. Notice that

$$\begin{aligned} p_{i+1} &= p_i^3(1 - \epsilon) + 3p_i^2(1 - p_i)(1 - \epsilon) + 3p_i(1 - p_i)^2\epsilon + (1 - p_i)^3\epsilon \\ &= (3p_i^2 - 2p_i^3)(1 - 2\epsilon) + \epsilon. \end{aligned}$$

Also, let $R(p_i) := p_{i+1}$, and let $\ell^*(\epsilon)$ be the largest real number such that $R(\ell^*(\epsilon)) = \ell^*(\epsilon)$. Note that $\ell^*(\epsilon)$ only exists when $\epsilon < 1/6$. Therefore for the following we assume that $\epsilon < 1/6$.

Lemma 34. It holds that $\ell^*(\epsilon) = \frac{1}{2} + \frac{1}{2} \sqrt{\frac{1-6\epsilon}{1-2\epsilon}}$. Furthermore, if $1/2 \leq p_i \leq \ell^*(\epsilon)$ then $p_i \leq p_{i+1} \leq \ell^*(\epsilon)$, and if $p_i \geq \ell^*(\epsilon)$ then $p_i \geq p_{i+1} \geq \ell^*(\epsilon)$, for all $i \in \{1, 2, \dots, \log_3 n\}$.

Proof. For any fixed point r of R , we have that $R(r) = r$ which implies that the fixed points are the zeros of the third order polynomial $R(r) - r$. Observe that these zeros are $\frac{1}{2} - \frac{1}{2}\sqrt{\frac{1-6\epsilon}{1-2\epsilon}}$, $\frac{1}{2}$, and $\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1-6\epsilon}{1-2\epsilon}}$.

For the second statement of the lemma, assume first that $p_i \geq \ell^*(\epsilon)$, and note that $R(p_i)$ is increasing because $R'(p_i) = (12\epsilon - 6)(p_i^2 - p_i) > 0$ for $\epsilon < 1/2$. Therefore we have

$$\ell^*(\epsilon) = R(\ell^*(\epsilon)) \leq R(p_i) = p_{i+1}.$$

Now we have to show that $p_{i+1} \leq p_i$. It is easy to see that $R(p_i) - p_i$ is decreasing when $p_i = \ell^*(\epsilon)$ and $\epsilon < 1/6$. Furthermore, $R(p_i) - p_i$ is a concave function when $p_i \geq 1/2$, since $(R(p_i) - p_i)'' = R''(p_i) = (12\epsilon - 6)(2p_i - 1) < 0$, for $p_i \geq 1/2$ and $\epsilon < 1/2$. Since $R(\ell^*(\epsilon)) - \ell^*(\epsilon) = 0$, it follows that for $p_i \in (\ell^*(\epsilon), 1)$, $R(p_i) - p_i \leq 0$, and thus $R(p_i) = p_{i+1} \leq p_i$.

The case $p_i \leq \ell^*(\epsilon)$ follows from the fact that $R(p_i) - p_i$ is decreasing when $p_i = \ell^*(\epsilon)$, $R(p_i) - p_i$ cannot be zero in the interval $(1/2, \ell^*(\epsilon))$, and $R(\ell^*(\epsilon)) - \ell^*(\epsilon) = 0$. \square

Note that the above technical lemma implies that $\ell^*(\epsilon)$ is a *stable* fixed point. The next two lemmas will be useful for setting the failure probabilities and analyzing the upper part of the tree.

Lemma 35. *Let G be a majority gate with input gates g_1, g_2 and g_3 which output 1 with probability $q_1 > 1/2, q_2 > 1/2$, and $q_3 > 1/2$, respectively. Furthermore let q_G be the probability that G outputs 1 (for the given probabilities of the inputs to output 1). If we alter g_1, g_2 , and g_3 to have probabilities $q'_1 > q_1, q'_2 > q_2$, and $q'_3 > q_3$ of outputting 1, then for the new probability q'_G of G outputting 1 it holds that $q'_G \geq q_G$.*

Proof. We have:

$$\begin{aligned} q'_G &= q'_1 q'_2 q'_3 (1 - \epsilon) + q'_1 q'_2 (1 - q'_3) (1 - \epsilon) + q'_1 q'_3 (1 - q'_2) (1 - \epsilon) \\ &\quad + q'_2 q'_3 (1 - q'_1) (1 - \epsilon) + q'_1 (1 - q'_2) (1 - q'_3) \epsilon + q'_2 (1 - q'_1) (1 - q'_3) \epsilon \\ &\quad + q'_3 (1 - q'_1) (1 - q'_2) \epsilon + (1 - q'_1) (1 - q'_2) (1 - q'_3) \epsilon = \\ &\quad (1 - 2\epsilon)(q'_1 q'_2 + q'_1 q'_3 + q'_2 q'_3 - 2q'_1 q'_2 q'_3) + \epsilon \geq \\ &\quad (1 - 2\epsilon)(q_1 q_2 + q_1 q_3 + q_2 q_3 - 2q_1 q_2 q_3) + \epsilon = q_G \end{aligned}$$

The inequality holds because the partial derivatives of the right-hand side with respect to q_1, q_2 and q_3 are all always nonnegative. \square

Lemma 36. *Consider a majority tree T of height $\lfloor 1/\delta \rfloor$ (for δ small enough) where each input of T is 1 with probability at least 0.79, and suppose that each gate of T has a failure probability of $\delta/10$. Then T outputs 1 with probability at least $1 - \delta$.*

Proof. By Lemma 35, the probability that T outputs 1 is minimized when all of the inputs are 1 with probability exactly 0.79.

Let $f(p_i) = p_{i+1} - p_i = p_i^3(1 - \epsilon) + 3p_i^2(1 - p_i)(1 - \epsilon) + 3p_i(1 - p_i)^2\epsilon + (1 - p_i)^3\epsilon - p_i$. Since by assumption $\epsilon < 1/2$ and $p_i > 1/2$, we have that $f''(p_i) = (12p_i - 6)(2\epsilon - 1)$ is negative and therefore f is concave. Furthermore, by setting $\epsilon = \delta/10$, and for δ small enough we may omit higher order terms and obtain

$$f(1 - \delta) \geq \frac{4}{5}\delta.$$

It can be verified that for δ small enough $f(0.79) \geq (4/5)\delta$, and $f(0.79) \geq f(1 - \delta)$. By the concavity of f , $f(x) \geq (4/5)\delta$ for $x \in (0.79, 1 - \delta)$. This means that there exists a $0 < k < 1/\delta$, so that each gate of height $\lfloor \log_3 n - 1/\delta + k \rfloor$ outputs 1 with probability at least $1 - \delta$. It suffices to show that each gate of height greater than $\lfloor \log_3 n - 1/\delta + k \rfloor$ has a probability of outputting 1 that is not lower than the probability of the gates one level below to output 1. This follows by Lemma 34: For δ small enough, $1 - \delta < \ell^*(\delta/10)$, and $p_i \leq p_{i+1}$. This completes the proof of the lemma. \square

With the help of the above lemmas, we are now ready to bound E_1 and E_2 .

Lemma 37. *It holds that $E_1(\delta) = \Omega(n \cdot E(\delta))$.*

Proof. In order to lower bound E_1 , we will consider the case where each input bit is 1. Note that the root cannot have a probability of failure greater than δ , since then even with all its inputs being correct it would not give the right output with the desired probability, and by Lemma 35 this probability can only decrease as the probability that the input gates are correct decreases. Because all gates must have the same probability, we have that each of the $O(n)$ gates has an energy consumption of at least $E(\delta)$, and the lemma follows. \square

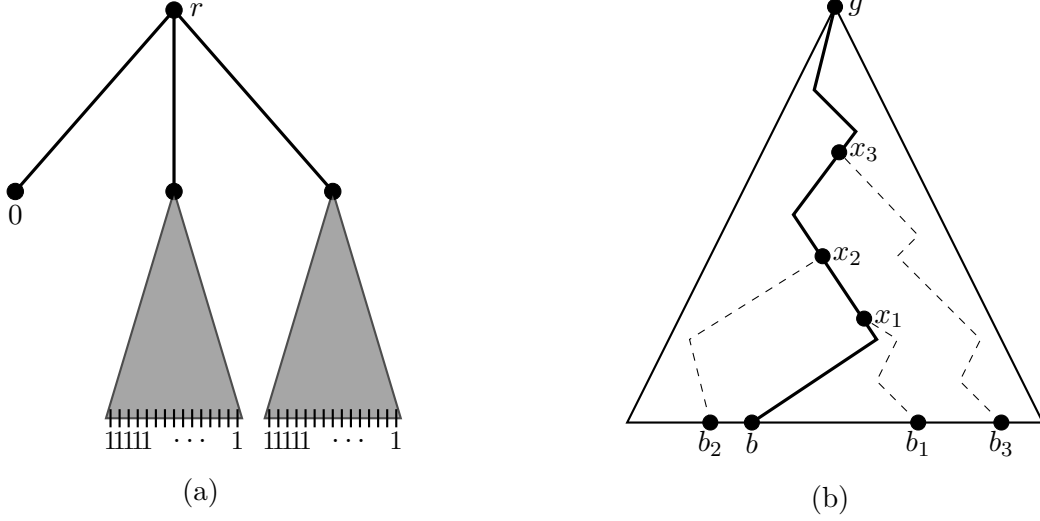


Figure 3: (a) $\Pr[r \text{ outputs } 1] \geq 1 - p$. (b) The path from b to g . The input gates b_i receive input 0.

Lemma 38. *It holds that $E_2(\delta) = O(n + 3^{1/\delta} E(\delta/10))$.*

Proof. Assume without loss of generality that the input contains at least $n - (1/2) \log_3 n$ 1's, so that the desired output is 1. We assign a failure probability of $\delta/10$ to each gate located at a height of at least $\log_3 n - 1/\delta$, and a failure probability of 0.12 to each gate at a height strictly less than $\log_3 n - 1/\delta$. By Lemma 36, it suffices to show that each gate at height $\lfloor \log_3 n - 1/\delta \rfloor$ outputs 1 independently with probability at least 0.79.

Let $p = 0.36$. Consider a majority tree where each gate has a failure probability of 0.12. Then, by Lemma 34, and since for this tree $p_0 = 0.88 > \ell^*(0.12)$, we have that the root of the tree outputs 1 with probability at least $\ell^*(0.12) > 0.8$. Thus, a reliable majority gate whose inputs are one 0 and the outputs of two arbitrarily sized majority trees whose inputs are all 1's outputs 0 with probability at most $1 - 0.8^2 = p$. See Figure 3a.

Consider a majority tree of height h rooted at gate g , and fix an input to this tree that contains exactly d zeros as input, with $0 < d < h$. Let b be any of the input gates of the tree that was assigned a zero for this input. We first show that the probability that the path

from b to gate g contains only 0's after each gate has computed is at most p^{h-d} . Let b_i for $i = 1, 2, \dots, d-1$ be the other input gates that were assigned 0's. We may assume that the path from each b_i to g intersects the path from b to g , at a distinct gate x_i . Furthermore we may assume that each such x_i outputs a 0. See Figure 3b for an example. The probability of such a path from b to g to contain only 0's is equal to the probability that the $h-d-1$ non- x_i gates on the path from b to g output a 0. Note that these non- x_i gates either receive a 0 and two inputs from majority subtrees whose inputs are all 1, or three inputs from majority subtrees whose inputs are all 1. Therefore, by the above observation about p and Lemma 35, the probability of such a path of all 0's is at most p^{h-d} .

Let T be any full (but not necessarily complete) majority tree of some height h_T . For any $h_A \geq h_T$, we can “complete” a copy of tree T by adding extra gates in order to obtain a perfect majority tree A of height h_A . We associate each gate in T with the corresponding gate in A . We claim that if the input at the leaves of both T and A consists of only 1's, then each gate of T is at least as likely to output 1 as the corresponding gate in A . We prove this claim by induction over the heights of gates in T . The base case, i.e., if a gate in T is a leaf, is straightforward. Assume now that each gate of T up to some height h' , has a higher probability of outputting 1 than its corresponding gate in A , and consider a non-leaf gate g' of T at height $h' + 1$. Since T is a full tree, and g' is not a leaf, g' must have three children. The inductive step now directly follows from Lemma 35.

Next, consider any subtree B of our original majority tree that is rooted at a gate g of height $\lfloor \log_3 n - 1/\delta \rfloor$. Since we assumed that the input to the original tree contains at most $(1/2) \log_3 n$ 0's, clearly this holds for B as well. See Figure 4 for an example of a tree B .

Now we want to lower bound the probability that g outputs a 1 when there is no path of all 0's from a leaf to g . We note that if there is no path of all 0's from a leaf to g then there exists a full subtree T' of B that is also rooted at g and whose inputs can be assumed to be all 1's. The subtree T' can be constructed by truncating each leaf-to-root path in B at the first node that outputs 1. The existence of T' follows from the fact that there is no path of all 0's from a leaf in B to g , but the structure T' depends on the random events occurring at each gate in $B \setminus T'$ and the leaves of T' . Conditioning on those random events, we have that B and T' output a 1 with the same probability.

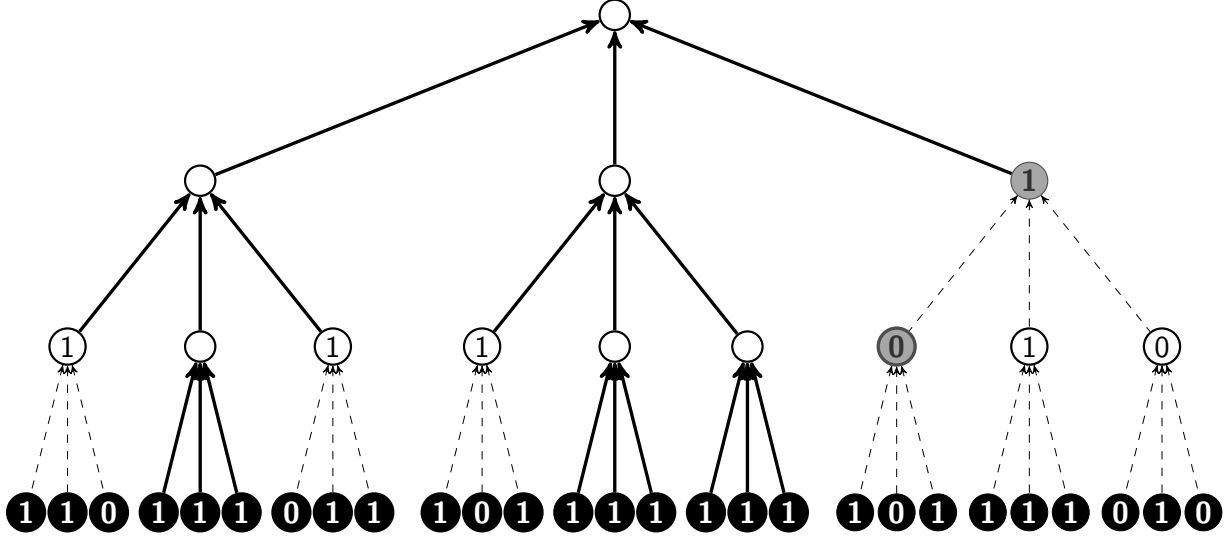


Figure 4: A subtree B . The solid edges denote the full ternary subtree $T \in \Gamma$. Note that T has 1's as inputs on its leaves. The dashed edges denote the edges in $B \setminus T$. The gray nodes denote gates that failed.

Let Γ be the set of all full ternary trees of height at most $\lfloor \log_3 n - 1/\delta \rfloor$, and for $T \in \Gamma$ let X_T be the event that the truncated tree described above is T . We have that:

$$\begin{aligned} \Pr[B \text{ outputs } 1] &\geq \\ \Pr[B \text{ outputs } 1 | \nexists \text{ path of all 0's}] \Pr[\nexists \text{ path of all 0's}] &\geq \\ \Pr[\nexists \text{ path of all 0's}] \Pr[A \text{ outputs } 1]. \end{aligned}$$

The second inequality follows because

$$\begin{aligned} \Pr[B \text{ outputs } 1 | \nexists \text{ path of all 0's}] &= \\ \sum_{T \in \Gamma} \Pr[X_T] \Pr[T \text{ outputs } 1 \text{ when given only 1's as input}] &\geq \\ \sum_{T \in \Gamma} \Pr[X_T] \Pr[A \text{ outputs } 1] &= \\ \Pr[A \text{ outputs } 1]. \end{aligned}$$

It follows by the union bound over all possible leaf-to-root paths of all 0's that g , and therefore every gate of height $\lfloor \log_3 n - 1/\delta \rfloor$, outputs 1 independently with probability at least $\ell^*(\epsilon) \cdot (1 - \frac{1}{2}(\log_3 n)p^{\frac{1}{2}\log_3 n - \frac{1}{\delta}})$. For n large enough this is at least 0.79. By Lemma 36, the upper part of the majority tree outputs 1 with probability at least $1 - \delta$. The total energy of the circuit is at most $E(0.12)n + 3^{1/\delta}E(\delta/10)$. \square

Proof. By Lemmas 37 and 38, we have that for $\delta' = \frac{1}{(n-c)\log_3 n}$,

$$\begin{aligned} \frac{E_1(\delta')}{E_2(\delta')} &= \Omega\left(\frac{n \cdot E(\delta')}{n + 3^{1/\delta'}E(\delta'/10)}\right) \\ &= \Omega\left(\frac{n \cdot E\left(\frac{1}{(1-c)\log n}\right)}{n + n^{1-c}E\left(\frac{1}{(1-c)\log n}\right)}\right) \\ &= \omega(1), \end{aligned}$$

where the second equality follows by the second property of easy-going functions, and the third equality by the first property of easy-going functions when taking n large enough. \square

We note that there are more trivial examples where heterogeneous supply voltages help. For example, consider a circuit that is a balanced binary tree of gates that each output the first bit, and the relation that outputs the first input bit. As most gates in this circuit are irrelevant to computing the desired relation, one can get an asymptotic energy saving by setting the supply voltages of the irrelevant gates to zero. Our example is more natural as one cannot simply power-off most of the gates. Although one might argue that our example is still not fully satisfactory as a more energy-efficient way to compute the super-majority relation is to use the majority circuit from [29] to compute the majority of the first $\log n$ bits with the supply voltages on each gate set so that the probability that any gate fails is at most δ . So a natural question is, “For every relation, is there is an asymptotically energy-optimal circuit for computing this relation that uses homogeneous supply voltages?”

5.0 HARDNESS AND ALGORITHMIC RESULTS FOR CIRCUIT ENERGY PROBLEMS

In this chapter, we evaluate the traditional solution to MCE, and show complexity-theoretic barriers to obtaining a better solution to MCE in general. The model of circuit and gate failure described in Chapter 2 is called the *von Neumann failure model*. In this chapter, partially to provide evidence that our results also hold for general models, we also prove results for the *0-default failure model*. In the 0-default failure model, gates are always faultless, but each input wire to a gate g is associated with a probability of failure ϵ , and when a wire fails it sends the default value of 0 (e.g., the wire by default carries a low voltage). More formally, for a given input $I = (b_1, \dots, b_{n_g}) \in \{0, 1\}^{n_g}$, the i^{th} input wire carries bit b_i . If $b_i = 0$ then with probability 1 g receives 0 as the i^{th} input bit. If $b_i = 1$, then with probability ϵ the wire fails and g receives 0 as the i^{th} input bit, and with probability $1 - \epsilon$ g receives 1 as the i^{th} input bit (note that a failure can only change a wire from carrying a 1 to carrying a 0). In this case, there is a voltage-to-energy function $P(v)$ mapping the supply voltage to the energy used by a wire with that supply voltage. The energy required by a circuit C is simply the aggregate energy used by the wires, $\sum_{w \in C} P(v)$. For convenience, we define a failure-to-energy function $E(q) := P(\epsilon^{-1}(q))$, where ϵ^{-1} denotes the inverse of the function ϵ . Thus the energy of a circuit C can be rewritten as $\sum_{w \in C} E(\epsilon(v))$. Since the two quantities we are most interested in are failure probability and energy, and the failure-to-energy function describes a direct relationship between the two, from henceforth we drop all reference to the supply voltage (e.g., we denote $\epsilon(v)$ by ϵ). (ϵ, δ) -reliability for the 0-default model is defined in the natural (analogous) way.

We consider bi-criteria approximations on energy and circuit failure.

Definition 39. For any circuit C and $\delta \in (0, 1)$, let $\epsilon_{C,\delta}^*$ be the solution to $MCE(C, \delta)$. An algorithm is a (c, d) -approximation for MCE if on any input (C, δ) it outputs a value ϵ such that C is $(\epsilon, d\delta)$ -reliable and $E(\epsilon) \leq c \cdot E(\epsilon_{C,\delta}^*)$.

Note that a $(c, 1)$ -approximation for MCE means that the approximation is only on energy, i.e., the algorithm outputs an ϵ such that the circuit is (ϵ, δ) -reliable and the circuit uses at most c times the energy of the circuit with the optimal choice of ϵ . Throughout this chapter we generalize our failure-to-energy function to be $E(\epsilon) = \Theta(\log^\alpha 1/\epsilon)$ for some $\alpha > 0$.

5.1 POLYNOMIAL-TIME APPROXIMATION OF THE MINIMUM CIRCUIT ENERGY PROBLEM

In this section we show in Theorem 40 that the approximation ratio of the traditional algorithm, which sets $\epsilon = \delta/s$, is $O(\log^\alpha s)$. We can actually prove a slightly more general bi-criteria approximation bound, in Theorem 41, that shows the trade-off on approximation between energy and reliability for a generalization of the traditional approach. For the 0-default failure model, we require that the circuit is *non-trivial* in the sense that there is at least one input that causes the output to be 0, and at least one input that causes the output to be 1.

Theorem 40. In both the von Neumann and 0-default failure models, the traditional approach is an $(O(\log^\alpha s), 1)$ -approximation for the MCE problem on non-trivial circuits.

Theorem 41. Let w denote the total number of wires of the circuit C , that is, $w = \sum_{g \in C} n_g$, and let φ denote the fan-in of the output gate of the circuit. In the 0-default failure model, setting $\epsilon = \delta/(\beta w)$, for any $\beta \geq 1$, yields a $((2\varphi^2/\log 2)^\alpha \log^\alpha(\beta w), 3/(2\beta))$ -approximate solution for the MCE problem on non-trivial circuits. In the von Neumann failure model, setting $\epsilon = \delta/(\beta s)$, for any $\beta \geq 1$, yields a $((2/\log 2)^\alpha \log^\alpha(\beta s), 3/(2\beta))$ -approximate solution for the MCE problem.

Proof. We first prove Theorem 41 for the 0-default failure model. First, we will choose the

greatest value of ϵ for which we can prove that the desired bound on the error of the circuit (that is, $3/(2\beta)$) is satisfied. Since the probability that no wire in the circuit C fails is $(1 - \epsilon)^w$, it is sufficient to set ϵ such that

$$(1 - \epsilon)^w \geq 1 - \frac{3}{2\beta}\delta,$$

that is

$$\log(1 - \epsilon) \geq \frac{\log\left(1 - \frac{3}{2\beta}\delta\right)}{w}. \quad (5.1)$$

From standard calculus we know that

$$\log(1 - x) > -\frac{3}{2}x \quad \text{for } 0 < x \leq 0.5828$$

and

$$\log(1 - x) < -x \quad \text{for } x < 1 \text{ and } x \neq 0,$$

and thus Inequality 5.1 is satisfied by setting $\epsilon = \delta/\beta w$, since

$$\log(1 - \epsilon) = \log\left(1 - \frac{\delta}{\beta w}\right) > -\frac{3}{2}\frac{\delta}{\beta w} > \frac{\log\left(1 - \frac{3}{2\beta}\delta\right)}{w}.$$

Then, we have to show that with this choice of ϵ the energy E used by the circuit is at most a factor of $(2\varphi^2/\log 2)^\alpha \log^\alpha(\beta w)$ of the energy E^* used in an optimal solution. As for the preceding theorem, to do this we determine an upper bound to the optimal solution ϵ^* , that is the maximum value of ϵ for which the circuit is (ϵ, δ) -reliable, from which it follows a lower bound for the energy used in an optimal solution. We have two cases, depending on whether the last gate g_o of the circuit outputs 0 or 1 on input $(0, 0, \dots, 0)$. Consider first the case, that is, $g_o(0, 0, \dots, 0) = 0$. The other case is symmetric. Since by hypothesis the circuit is non-trivial, then the circuit does not represent the constant function $f' = 0$. Hence, there must be at least one input I to the circuit C for which $C(I) = 1$. Let q denote the probability that all the φ wires entering the output gate g_o receive value 0 when the input

to the circuit is I . If we denote with p the probability that the circuit outputs the correct bit when each of its wires fails with probability ϵ , then it holds that

$$\begin{aligned}
1 - p &= \mathbf{Pr}[\text{circuit } C \text{ outputs the wrong bit}] \\
&\geq \mathbf{Pr}[\text{circuit } C \text{ outputs the wrong bit on input } I] \\
&= \mathbf{Pr}[\text{circuit } C \text{ outputs 0 on input } I] \\
&= q \cdot 1 + (1 - q) \cdot \\
&\quad \cdot \mathbf{Pr}[g_o \text{ receives an input } x \text{ s.t. } g_o(x) = 0] \\
&\geq q + (1 - q) \cdot \mathbf{Pr}[g_o \text{ receives input } x = (0, 0, \dots, 0)] \\
&\geq q + (1 - q) \mathbf{Pr}[\text{all the } \varphi \text{ input wires of gate } g_o \text{ fail}] \\
&= q + (1 - q)\epsilon^\varphi \\
&\geq \epsilon^\varphi,
\end{aligned}$$

and therefore,

$$p \leq 1 - \epsilon^\varphi.$$

In an optimal solution it must be that

$$p \geq 1 - \delta,$$

and thus, combining the two previous inequalities, it must hold that

$$1 - \delta \leq 1 - (\epsilon^*)^\varphi,$$

that is

$$(\epsilon^*) \leq \delta^{1/\varphi}.$$

This implies a lower bound of $s \log^\alpha(1/\delta^{1/\varphi})$ for the optimal energy consumption E^* .

For the same reason, the energy consumption E of our approximate solution is $s \log^\alpha(\beta w/\delta)$. Since $\delta < 1/2$, $\beta \geq 1$, we have,

$$\begin{aligned}
E &= n \log^\alpha \left(\frac{\beta w}{\delta} \right) \\
&= n \left(\log(\beta w) + \log \frac{1}{\delta} \right)^\alpha \\
&\leq n 2^{\alpha-1} \left(\log^\alpha(\beta w) + \log^\alpha \frac{1}{\delta} \right) \\
&= n 2^{\alpha-1} \left(\log^\alpha(\beta w) + \varphi^\alpha \log^\alpha \frac{1}{\delta^{1/\varphi}} \right) \\
&\leq n 2^{\alpha-1} \left(\log^\alpha(\beta w) \frac{\varphi^\alpha \log^\alpha \frac{1}{\delta^{1/\varphi}}}{\log^\alpha 2^{1/\varphi}} + \right. \\
&\quad \left. + \log^\alpha(\beta w) \frac{\varphi^\alpha \log^\alpha \frac{1}{\delta^{1/\varphi}}}{\log^\alpha 2^{1/\varphi}} \right) \\
&= n \left(\frac{2\varphi^2}{\log 2} \right)^\alpha \log^\alpha(\beta w) \log^\alpha \frac{1}{\delta^{1/\varphi}} \\
&\leq \left(\frac{2\varphi^2}{\log 2} \right)^\alpha \log^\alpha(\beta w) \cdot E^*,
\end{aligned}$$

where the first inequality follows from Jensen's inequality.

The proof for the von Neumann failure model is similar. □

We can then prove Theorem 40, showing that the traditional approach is a $(O(\log^\alpha s), 1)$ -approximation, by using the same analysis with $\beta = 3/2$,

5.2 HARDNESS OF APPROXIMATION FOR THE MINIMUM CIRCUIT ENERGY PROBLEM

In this section we prove that it is NP-hard to obtain a significantly better approximation than the $O(\log^\alpha s)$ obtained from the traditional approach.

Theorem 42. *In both the von Neumann and 0-default failure models, it is NP-hard to $(\log^{\alpha-\gamma} s, 1)$ -approximate the MCE problem for any $\gamma > 0$.*

Proof. The main idea of the proof is to show that for a satisfiable circuit and an unsatisfiable circuit there is a large gap between the probability they correctly compute their input. In particular, in the case of a satisfiable input, we show that it is very unlikely for the output of the circuit to be a 1. For technical reasons we restrict γ to $\gamma \in (0, \alpha)$. It is clear that the problem is only computationally harder as γ increases. The proof makes use of some technical facts stated after this proof.

Assume by contradiction that there exists a $(\log^{\alpha-\gamma} s, 1)$ -approximate algorithm A . For notational convenience, let $c = \log^{\alpha-\gamma} s$. Furthermore, let ϕ be an arbitrary 3SAT formula with n variables and m clauses. Let S_ϕ be the natural circuit for ϕ that uses at most $3m$ NOT gates to represent the negated variables, m OR gates of fan-in 3 to represent the clauses, and a tree of $m - 1$ AND gates of fan-in 2 that compute the conjunction of all clauses. See Figure 5 for an example.

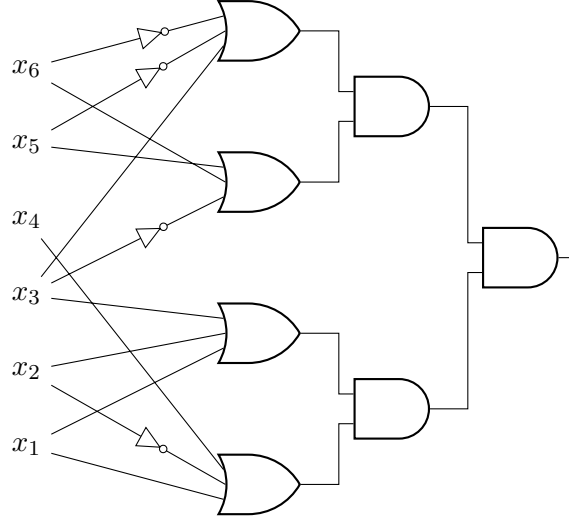


Figure 5: The circuit S_ϕ where $\phi = (x_1 \vee \bar{x}_2 \vee x_4) \wedge (x_1 \vee x_2 \vee x_3) \wedge (\bar{x}_3 \vee x_6 \vee x_5) \wedge (x_3 \vee \bar{x}_5 \vee \bar{x}_6)$.

Let $s = |S_\phi|$, and note that $2m - 1 \leq s \leq 5m - 1$. We choose ϵ such that

$$\left(1 - \epsilon^{\sqrt[3]{c}}\right)^{m+1} + 4\epsilon^{\sqrt[3]{c}} < 1 - 8\epsilon \quad (5.2)$$

and let $\delta = 8\epsilon$. We will show later that such an ϵ must exist. Now, consider the output of A, ϵ_A on S_ϕ with input δ . We claim that ϕ is satisfiable if and only if $E(\epsilon_A) > c \log^\alpha(\frac{1}{\epsilon})$. In

the first case, assume ϕ is satisfiable. Consider S_ϕ where each gate fails independently with probability ϵ_A (or wire in the 0-default model), and the input x such that $\phi(x) = 1$. Let E_0 be the event that each of the OR gates receives at least 1 positive input, E_1 be the event that all of the OR gates output a 1 and E_2 be the event that S_ϕ outputs a 1. By Lemma 45, we know that

$$\Pr[E_2] \leq (1 - \epsilon_A)^{m+1} + 4\epsilon_A.$$

Further by (5.2), if $\epsilon_A = \epsilon^{\sqrt[m]{c}}$ then we have

$$\Pr[E_2] \leq (1 - \epsilon^{\sqrt[m]{c}})^{m+1} + 4\epsilon^{\sqrt[m]{c}} < 1 - 8\epsilon = 1 - \delta. \quad (5.3)$$

Note that for $\epsilon_A \in [\epsilon^{\sqrt[m]{c}}, 1/2)$, the quantity $(1 - \epsilon_A)^{m+1} + 4\epsilon_A$ is maximized at $\epsilon_A = \epsilon^{\sqrt[m]{c}}$. Therefore, we must have $\epsilon_A < \epsilon^{\sqrt[m]{c}}$ otherwise by (5.3), the probability that S_ϕ is correct would not be within $1 - \delta$, contradicting that A is a $(\log^{\alpha-\gamma}(s), 1)$ -approximation. Further since E is decreasing, $E(\epsilon_A) > E(\epsilon^{\sqrt[m]{c}}) = c \log^\alpha(\frac{1}{\epsilon})$.

Now, assume that ϕ is unsatisfiable. Consider S_ϕ with an arbitrary input x where each gate fails independently with probability ϵ_A (or wire in the 0-default model). Note since ϕ is unsatisfiable, $\phi(x) = 0$. Let the events E_0, E_1 and E_2 be defined in the same way as before. Using the bounds on $\Pr[E_2|E_1]$ and $\Pr[E_2|\neg E_1]$ from the proof of Lemma 45 we have,

$$\begin{aligned} \Pr[E_2] &= \Pr[E_2|E_1] \Pr[E_1] + \Pr[E_2|\neg E_1] \Pr[\neg E_1] \\ &= (1 - \epsilon_A) \Pr[E_1] + (4\epsilon_A) \Pr[\neg E_1] \\ &= (1 - \epsilon_A)(\Pr[E_1|E_0] \Pr[E_0] \\ &\quad + \Pr[E_1|\neg E_0] \Pr[\neg E_0]) + (4\epsilon_A) \Pr[\neg E_1] \\ &\leq (1 - \epsilon_A) \left((1 - \epsilon_A)^m (3\epsilon_A) + (1 - \epsilon)^{m-1} (\epsilon_A) \cdot 1 \right) \\ &\quad + (4\epsilon_A) \Pr[\neg E_1] \\ &\leq 8\epsilon_A. \end{aligned}$$

Therefore for all inputs, the probability that S_ϕ is correct is at least $1 - 8\epsilon_A$. So note that if $\epsilon_A = \epsilon$, the probability S_ϕ is correct is at least $1 - \delta$. This shows that $\epsilon^* \geq \epsilon$. By

the definition of A being $(c, 1)$ approximate this means that $E(\epsilon_A) \leq cE(\epsilon) = c\log^\alpha(\frac{1}{\epsilon})$. This shows that we can determine the satisfiability of ϕ using A . If $E(\epsilon_A) > c\log^\alpha(\frac{1}{\epsilon})$, ϕ is satisfiable, and otherwise if $E(\epsilon_A) \leq c\log^\alpha(\frac{1}{\epsilon})$, ϕ is not satisfiable. The last thing to do is show the existence of an ϵ satisfying (5.2). Consider $\epsilon = (\frac{1}{m+1})^{\frac{1}{\sqrt[c]{c}}}$. Then,

$$(1 - \epsilon^{\sqrt[c]{c}})^{m+1} + 4\epsilon^{\sqrt[c]{c}} \leq e^{-\epsilon^{\sqrt[c]{c}}(m+1)} + 4\epsilon^{\sqrt[c]{c}} = e^{-1} + \frac{4}{m+1}.$$

Also, $1 - 8\epsilon = 1 - 8(\frac{1}{m+1})^{\frac{1}{\sqrt[c]{c}}}$. Note that $e^{-1} + \frac{4}{m+1} < 1 - 8(\frac{1}{m+1})^{\frac{1}{\sqrt[c]{c}}}$ since

$$\lim_{m \rightarrow \infty} 8 \left(\frac{1}{m+1} \right)^{\frac{1}{\sqrt[c]{c}}} \leq \lim_{m \rightarrow \infty} 8 \left(\frac{1}{m+1} \right)^{\frac{1}{\log^{1-\frac{\gamma}{\alpha}} m}} \rightarrow 0.$$

□

We now state and prove some technical lemmas used in the above proof.

Lemma 43. *The recurrence $p_i = p_{i-1}^2(1 - \epsilon) + (1 - p_{i-1}^2)\epsilon$, $p_0 = 1$ satisfies $p_i \leq p_{i-1}$ for all i .*

Proof. By taking the derivative of p_i with respect to p_{i-1} , we see it is increasing in p_{i-1} and therefore $p_i \leq p_{i-1}$ implies $p_{i+1} \leq p_i$. Since $p_1 < p_0$, combining these facts gives that $p_i \leq p_{i-1}$ for all i . □

Lemma 44. *Let $\epsilon = (1/(m+1))^{1/\sqrt[\alpha]{\log^{\alpha-\gamma}(s)}}$ and let $p_i = p_{i-1}^2(1 - \epsilon) + (1 - p_{i-1}^2)\epsilon$, with $p_0 = 1$. Then, for m bigger than some constant M_0 ,*

$$p_{\log_2 m} \leq 3\epsilon.$$

Proof. We first show that $p_{\log_2(1/\epsilon)} \leq 1/\sqrt{2}$. Note that for $p_{i-1} \geq 1/\sqrt{2}$, $p_i = p_{i-1}^2(1 - 2\epsilon) + \epsilon = p_{i-1}^2 - \epsilon(2p_{i-1}^2 - 1) \leq p_{i-1}^2$. Therefore, since $p_1 = (1 - \epsilon)$ it follows that for $p_i \geq 1/\sqrt{2}$, $p_i \leq (1 - \epsilon)^{2^i}$. Note for $i = \log_2(1/\epsilon)$ we have $p_i \leq (1 - \epsilon)^{1/\epsilon} \rightarrow 1/e$ as $\epsilon \rightarrow 0$. It follows that for some constant M_0 and $m \geq M_0$, $p_{\log_2(1/\epsilon)} \leq 1/\sqrt{2}$. Further it is easy to see that $p_{\log_2(1/\epsilon)+3} \leq 1/8$ by expanding the recurrence and letting $\epsilon \rightarrow 0$.

The last thing to show is that in $\log_2 \log_2(1/\epsilon)$ additional steps we can go from $1/8$ to 3ϵ . Let $k = \log_2(1/\epsilon) + 3$. Then, this follows by noting that $p_{k+j} \leq p_k^{2^j} + 2\epsilon$ for any $j \geq 0$. To

see this note that it clearly holds for $j = 0$, and by induction if this holds for an arbitrary $j > 0$, then

$$p_{k+j+1} = p_{k+j}^2(1 - 2\epsilon) + \epsilon \leq (p_k^{2^j} + 2\epsilon)^2 + \epsilon \leq p_0^{2^{j+1}} + 3\epsilon.$$

By solving $p_k^{2^j} = \epsilon$ we obtain that $p_{\log_2(1/\epsilon)+3+\log_2 \log_2(1/\epsilon)} \leq 3\epsilon$. Lastly, note that

$$\begin{aligned} \log_2(1/\epsilon) + \log_2 \log_2(1/\epsilon) + 3 &\leq 3 \log_2(1/\epsilon) \\ &= 3 \frac{\log_2(m+1)}{\sqrt[\alpha]{\log^{\alpha-\gamma} s}} \\ &\leq 3 \frac{\log_2(m+1)}{\sqrt[\alpha]{\log^{\alpha-\gamma} m}} \sqrt[\alpha]{\log^{\alpha-\gamma} e} \\ &\leq 9 \log_2^{\gamma/\alpha}(m) \end{aligned}$$

and so $p_{\log_2 m} \leq 3\epsilon$, since by Lemma 43 $p_{\log_2 m} \leq p_{9 \log^{\gamma/\alpha}(m) / \sqrt[\alpha]{\log^{\alpha-\gamma} 2}}$. \square

Lemma 45. *Let ϕ be some satisfiable 3SAT formula with n variables and x be the input such that $\phi(x) = 1$. Then, in both the von Neumann model and the 0-default model, the probability that S_ϕ outputs a 1 is bounded above by $(1 - \epsilon)^{m+1} + 4\epsilon$, where $\epsilon = (1/(m+1))^{1/\sqrt[\alpha]{\log^{\alpha-\gamma}(s)}}$.*

Proof. We first show this holds in the von Neumann failure model. Let g_o be the output gate of C (the root of the tree of AND gates). Further, let E_0 be the event that each of the OR gates receives at least 1 positive input, E_1 be the event that all of the OR gates output a 1 and E_2 be the event that g_o outputs a 1. We first calculate $\Pr[E_2|E_1]$. This is the probability that the tree of $m - 1$ AND gates outputs a 1 when all the inputs to the leaves are 1. Let p_i be the probability that a gate on the i^{th} level outputs a 1. We define the input to the leaves to be at level 0. Note that $p_0 = 1$, and for $i > 0$, we can write p_i as a recurrence in the form,

$$p_i = p_{i-1}^2(1 - \epsilon) + (1 - p_{i-1}^2)\epsilon.$$

Further, since $p_1 = (1 - \epsilon)$, and by Lemma 43, the sequence p_i is decreasing as $i \rightarrow \infty$ we have that $\Pr[E_2|E_1] \leq (1 - \epsilon)$. Next, we bound $\Pr[E_2|\neg E_1]$. Let A denote the event that g_o receives two 1's as input. We have,

$$\begin{aligned}
\mathbf{Pr}[E_2|\neg E_1] &= \mathbf{Pr}[E_2|\neg E_1 \wedge A] \mathbf{Pr}[A|\neg E_1] \\
&\quad + \mathbf{Pr}[E_2|\neg E_1 \wedge \neg A] \mathbf{Pr}[\neg A|\neg E_1] \\
&\leq (1 - \epsilon) \mathbf{Pr}[A|\neg E_1] + \epsilon.
\end{aligned}$$

The last thing to do is bound $\mathbf{Pr}[A|\neg E_1]$. Informally, we first argue that the probability of getting a 1 to the root of the tree is only increased if E_1 occurs, that is all leaves have value 1. After that, we can use the recurrence to show that for sufficiently large trees this probability is $O(\epsilon)$. More formally, for some fixed gate g' , let p_L be the probability the left input is 1 and p_R be the probability the right input is 1. Then, if $p_{g'}$ denotes the probability g' outputs a 1, we have

$$p_{g'} = (p_L p_R)(1 - \epsilon) + (1 - p_L p_R)\epsilon.$$

Taking the partial derivative with respect to p_L or p_R shows that $p_{g'}$ will increase as p_L or p_R increase. This implies that $\mathbf{Pr}[A|\neg E_1] \leq \mathbf{Pr}[A|E_1]$, since for every leaf, the probability of having a 1 will not decrease, and therefore by induction on the levels of the tree, every gate will have an increased probability of outputting a 1. Let h be the height of the tree. Then, note that $\mathbf{Pr}[A|E_1] = p_h^2 \leq p_h$ as defined by the recurrence in Lemma 43. However, since $h = \log_2 m$ by Lemma 44 $p_{\log_2 m} \leq 3\epsilon$ and therefore $\mathbf{Pr}[A|\neg E_1] \leq 3\epsilon$ and further, $\mathbf{Pr}[E_2|\neg E_1] \leq 4\epsilon$. We are now ready to calculate the probability that $S_\phi(x)$ outputs a 1. We have,

$$\begin{aligned}
\mathbf{Pr}[E_2] &= \mathbf{Pr}[E_2|E_1] \mathbf{Pr}[E_1] + \mathbf{Pr}[E_2|\neg E_1] \mathbf{Pr}[\neg E_1] \\
&= (1 - \epsilon) \mathbf{Pr}[E_1] + 4\epsilon \mathbf{Pr}[\neg E_1] \\
&= (1 - \epsilon)(\mathbf{Pr}[E_1|E_0] \mathbf{Pr}[E_0] \\
&\quad + \mathbf{Pr}[E_1|\neg E_0] \mathbf{Pr}[\neg E_0]) + 4\epsilon \mathbf{Pr}[\neg E_1] \\
&\leq (1 - \epsilon) \left((1 - \epsilon)^m \cdot 1 + (1 - \epsilon)^{m-1}(\epsilon) \cdot 1 \right) + 4\epsilon \\
&\leq (1 - \epsilon)^{m+1} + 4\epsilon.
\end{aligned}$$

To see that this holds in the 0-default model, note that $\Pr[E_2|\neg E_1] = 0 \leq 4\epsilon$ since a 0 wire will never flip to a 1. Using this we can make an identical calculation to the above to get that $\Pr[E_2] \leq (1 - \epsilon)^{m+1} + 4\epsilon$. \square

We end by noting that a slight modification of the proof of Theorem 42 can be used to prove the following more general theorem.

Theorem 46. *It is NP-hard to (c, d) -approximate the MCE problem in both the von Neumann and 0-default failure models for all $c > 1$ and d such that $\lim_{m \rightarrow \infty} 8d \left(\frac{1}{m+1}\right)^{\frac{1}{\sqrt[c]{c}}} \rightarrow 0$.*

5.3 HARDNESS OF DETERMINING (ϵ, δ) -RELIABILITY ON FIXED INPUTS

In this section we prove the following theorem.

Theorem 47. *In the 0-default failure model, given ϵ, δ, C , and I , it is NP-Hard to determine if C is (ϵ, δ) -reliable on I .*

The section proceeds as follows. Our reduction is from the gap-3SAT problem, which is known to be NP-Hard for certain parameters, so we begin by formally defining this problem. We then bound the probability that the natural 3SAT circuit, S_ϕ , outputs a 1 when given a random input both when ϕ is satisfiable, and when at most 15/16 fraction of the clauses of ϕ are satisfiable. Finally, we introduce a circuit N_k that, in the presence of failures, can be used to randomize our input.

First we must introduce the gap-3SAT $[\alpha, \beta]$ problem (with $\alpha \leq \beta$), as the NP-hardness reduction will be from this problem. The problem is as follows: Given a 3SAT instance, output “YES” if at least a β fraction of the clauses are satisfiable, “NO” if at most an α fraction of the clauses are satisfiable, and either “YES” or “NO” otherwise (i.e., such inputs are not given). The hardness of this problem for certain values of α and β follows from the PCP Theorem [7], and in particular, Håstad proved the following theorem, giving the best possible values for α and β .

Theorem 48 (Håstad [19]). *Gap-3SAT $[7/8 + \epsilon, 1]$ is NP-Hard for all $\epsilon > 0$.*

The reduction is from the hardness of gap-3SAT[15/16,1]. We use as our main circuit the standard 3SAT circuit S_ϕ used elsewhere in this paper (see Figure 5 and the related discussion). As we have seen, if the tree of AND gates does not receive all 1's, then with probability 1 the output is 0. Thus, intuitively, if we could give S_ϕ a random input, then (i) if ϕ is satisfiable, on the satisfying input S_ϕ is much more likely to output a 1 than on any other input, and (ii) if ϕ is not satisfiable, then any assignment satisfies a fraction of at most 15/16 of the clauses, so a large number (for example, at least a 1/16 fraction) of wires would have to fail for S_ϕ to be likely to output a 1. We first bound the probability that S_ϕ outputs a 1 when receiving an almost random input in the two cases when there exists a satisfying assignment and when at most a 15/16 fraction of the clauses can be satisfied. We then show that it is possible with a polynomially sized circuit to create an almost random input from a fixed input, and use this to complete the reduction.

Lemma 49. *Let ϕ be a 3SAT formula and S_ϕ be the circuit for ϕ , where each wire fails independently with probability ϵ . Suppose that each input to S_ϕ is a 1 with probability at least $1/2 - \gamma$ and at most $1/2 + \gamma$. Then, in the 0-default failure model:*

1. *If ϕ is satisfiable, then*

$$\Pr[S_\phi \text{ outputs a 1}] \geq \left(\frac{1}{2} - \gamma\right)^n (1 - \epsilon)^{5m}.$$

2. *If at most a 15/16 fraction of the clauses of ϕ are satisfiable, then*

$$\Pr[S_\phi \text{ outputs a 1}] \leq (3\epsilon)^{m/16}.$$

Proof. Let O be the random output of the circuit S_ϕ and A be the random event that the tree of AND gates of S_ϕ receives all 1's as input. Then clearly $O = 1$ if A occurs and none of the wires within the tree of AND gates fail, and $O = 0$ otherwise. Therefore,

$$\Pr[O = 1] = (1 - \epsilon)^{2m-1} \Pr[A].$$

1. **ϕ is satisfiable.** Let E be the event that S_ϕ receives a satisfying assignment as input. The probability E occurs is at least $(\frac{1}{2} - \gamma)^n$, since this is a lower bound on S_ϕ receiving any fixed input. Further, if none of the wires entering the OR gates fail (the wires entering NOT gates in the clauses can only fail and output 1, which only increases the probability that $O = 1$), then A occurs, so

$$\Pr[A|E \wedge \phi \text{ is satisfiable}] \geq (1 - \epsilon)^{3m}.$$

Clearly, the probability that A occurs if S_ϕ does not receive satisfying assignment as input is at least 0, so the first statement of the lemma follows since

$$\begin{aligned} \Pr[O = 1|\phi \text{ is satisfiable}] &\geq \\ (1 - \epsilon)^{2m-1} \Pr[A|\phi \text{ is satisfiable}] &\geq \\ (1 - \epsilon)^{2m-1} \Pr[E|\phi \text{ is satisfiable}] &\cdot \\ \Pr[A|E \wedge \phi \text{ is satisfiable}] &\geq \\ \left(\frac{1}{2} - \gamma\right) (1 - \epsilon)^{5m-1}. \end{aligned}$$

2. **At most a 15/16 fraction of the clauses of ϕ are satisfiable.** For this case, every assignment satisfies at most a 15/16 fraction of the clauses. Thus we have that an upper bound on A occurring is if at least one of the wires associated with not gates in every clause that is not satisfied fails (if a wire entering an OR gate fails the gate will output 0), and all other gates do not fail. Thus we have that

$$\Pr[A|\phi \text{ is not satisfiable}] \leq (3\epsilon)^{m/16},$$

and therefore

$$\Pr[O = 1|\phi \text{ is not satisfiable}] \leq (3\epsilon)^{m/16}.$$

□

The following circuit will be useful in the reduction.

Definition 50. N_k is the circuit consisting of one input bit connected to a single line of k NOT gates, i.e., the output of the i th NOT gate is the input to the $i + 1$ st NOT gate, for $i \in [k - 1]$.

If no gate in N_k fails, the output on input bit b is $(b + k) \bmod 2$. However, if each of these gates fail independently with probability ϵ , then the output is random and, for k large enough, will be b with probability very close to $\frac{1}{2}$. Consider the Markov chain M with two states that correspond to the output bit after a certain number of NOT gates, and transitions with probabilities based on whether or not the wire entering the current NOT gate fails. If we label one state “1” and the other “0”, then the output of N_k is identical to the output of starting M in state b and running for k steps. The transition from the 0 state to the 1 state happens with probability 1, since the wire cannot fail in this case. On the other hand, the transition from the 1 state to the 0 state only happens with probability $1 - \epsilon$, and the chain stays in the 1 state with probability ϵ . It is easy to verify that this chain is irreducible, aperiodic, and reversible. The transition matrix is

$$M = \begin{bmatrix} 0 & 1 - \epsilon \\ 1 & \epsilon \end{bmatrix}.$$

The eigenvalues of M are 1 and $\epsilon - 1$, and the stationary distribution of M is $\frac{1-\epsilon}{2-\epsilon}$ in state 0, and $\frac{1}{2-\epsilon}$ in state 1, so the number of steps $k(\rho)$ until we are ρ away from the stationary distribution is

$$k(\rho) \leq \frac{1}{\epsilon} \log \left(\frac{2 - \epsilon}{\rho(1 - \epsilon)} \right).$$

For a more in depth discussion of Markov chains and mixing times, see, e.g., [23]. By setting $\rho = 0.05$, we obtain the following observation.

Observation 51. *Suppose each wire of N_k fails independently with probability $\epsilon < 1/10$. Then in the 0-default failure model if $k \geq \log(44)/\epsilon$, we have that $0.4 \leq \Pr[N_k(b) = b] \leq 0.6$.*

We can now finish the reduction.

Proof of Theorem 47. The reduction is from gap-3SAT[15/16,1]. Let ϕ be a 3SAT formula that is either satisfiable or at most a 15/16 fraction of the clauses can be satisfied. Without loss of generality, we can assume the assignments of all 1’s and all 0’s do not satisfy ϕ , and

that there are at least n clauses in ϕ . We set $\epsilon = 1.4 \times 10^{-7}$ (a constant). Construct a circuit S'_ϕ that is S_ϕ except that each input first passes through a N_k circuit, where $k = \lceil \log(44)/\epsilon \rceil$, and thus S'_ϕ is polynomial in size and logarithmic in depth. We fix the input to this circuit to be the input of all 1's, so the correct output of S'_ϕ is 0. By Observation 51, the output of each N_k circuit is 1 with probability at least $\frac{1}{2} - \gamma$ and at most $\frac{1}{2} + \gamma$ for $\gamma = 0.1$. We set $\delta = (3\epsilon)^{m/16}$. By Lemma 49 (since S'_ϕ is incorrect if it outputs 1), if we show that

$$(3\epsilon)^{m/16} < (0.4)^n (1 - \epsilon)^{5m} \quad (5.4)$$

then it is NP-Hard to determine whether or not S'_ϕ outputs correctly with probability at least $1 - \delta$. Rearranging the exponents and noting that $n \leq m$, we obtain that

$$3\epsilon < (0.4)^{16} (1 - \epsilon)^{80}$$

implies that (5.4) holds. It is easy to verify that the choice of ϵ satisfies this inequality. \square

In the von Neumann failure model, we were unable to prove that determining if a circuit is (ϵ, δ) -reliable is NP-Hard. The difficulty with following this same proof structure extends from two conflicting constraints. The first constraint is that the probability that S'_ϕ outputs a 1 when at most a 15/16 fraction of the clauses of ϕ are satisfiable must be smaller than the probability S'_ϕ outputs a 1 when ϕ is satisfiable. In the 0-default failure model, if the tree of AND gates in S'_ϕ received anything but all 1's as input, the circuit would output 0. In the von Neumann failure model this is not the case, since any of the AND gates (e.g., the output gate) can fail and incorrectly output a 1 instead of a 0. Thus in the von Neumann failure model, the tree of AND gates has a higher probability of outputting a 1 if 15/16 of its inputs are 1's than if very few of its inputs are 1's, and this difference in probability is polynomial in ϵ . Since in the case when ϕ is satisfiable we can only guarantee that the output of the N_k circuits is the satisfying assignment with probability approximately 2^{-n} , we need to require ϵ to be exponentially small in order to guarantee that S'_ϕ has a higher probability of outputting a 1 when it is satisfiable than when at most a 15/16 fraction of the clauses are satisfiable. The second constraint is that we need $k \geq 1/\epsilon$ in order for N_k to output a random bit. Thus if ϵ is exponentially small, the circuit S'_ϕ will be exponentially large, and so the reduction will not be polynomial time.

5.4 TREE CIRCUITS

There are classes of circuits for which the problems discussed in this paper are much easier, namely circuits whose graph representation is a tree. The hardness results in this paper stem from the fact that, in general, the undirected version of the DAG representing a circuit C may contain cycles. When this is not the case, then the probability that a gate g outputs a 1 or a 0 is dependent only on the outcomes of the immediate predecessors of g in C , and thus the situation is much simpler. Given a circuit C that is a tree and where each gate has bounded fan-in, we describe below how to, in both the von Neumann and 0-default failure models, answer the question of whether C is (ϵ, δ) -reliable in time polynomial in the size of C (it can be seen that polynomial complexity can be achieved also in slightly more general settings, e.g., when the circuit's structure is “close to” a tree).

The algorithm is as follows: Each gate g stores four probabilities:

1. The highest probability that g is correct given that its correct output is 1.
2. The lowest probability that g is correct given that its correct output is 1.
3. The highest probability that g is correct given that its correct output is 0.
4. The lowest probability that g is correct given that its correct output is 0.

Let φ be the fan-in of g , and let g_1, \dots, g_φ be the parents of g . By choosing one of the stored probabilities from each of g 's parents, we can in $O(2^\varphi)$ steps calculate the probability that g outputs a 1 in that case, and the correct output of g in that case can be computed from the correct outputs for the probabilities chosen from g 's parents. Since there are 4^φ ways to choose one stored probability from each of g 's parents, we calculate all of these probabilities. Of those where the correct output of g is a 1, we find and store the highest and lowest probabilities that g does output a 1, and do the same for those where the correct output of g is a 0. At the output gate, we find the minimum of the lowest probability that g is correct given that its correct output is 1, and the lowest probability that g is correct given that its correct output is 0. This value determines the minimum value for δ given that functional failures occur with probability ϵ . It is straightforward to see how this algorithm could be modified slightly to find the input to the circuit that minimizes the probability of

correctness when functional failures occur with probability ϵ .

To see why this algorithm is correct, consider the situation where all but the i th parent, g_i , of some gate g output a 1 with fixed probability. In this case, the probability that g outputs a 1 is linear in the probability that g_i outputs a 1, and thus this probability is monotonically increasing, monotonically decreasing, or constant, as the probability that g_i outputs a 1 increases. Further, since the circuit is a tree, changing the input to the subtree rooted at g_i does not affect the probability that any of the other parents of g output a 1. Thus we can compute the highest and lowest probabilities that g will output a 1 by some combination of the highest and lowest probabilities that its parents will output a 1. Since we do not know what the correct output for g should be on the input that causes the circuit to be incorrect with highest probability, we store these probabilities in the cases when the correct output of g is either 1 or 0.

5.5 NON-MONOTONICITY OF δ IN ϵ

For any circuit C , let $\delta^*(\epsilon)$ be the smallest value such that C is $(\epsilon, \delta^*(\epsilon))$ -reliable. A question that one might ask is whether $\delta^*(\epsilon)$ is, in general, a non-decreasing function of ϵ . In both the von Neumann and 0-default failure models, this is not the case. The circuit depicted in Figure 6 provides an example for the von Neumann failure model. For this circuit, we obtain different bounds on $\delta^*(\epsilon)$ depending on the input to the circuit. The three cases are based

on how many of the OR gates receive a 1 as input.

$$\begin{aligned}
\delta^*(\epsilon) &\geq \Pr[y \neq (x_1 \vee x_2) \wedge (x_3 \vee x_4) | (x_1 \vee x_2) = 1 \\
&\quad \text{and } (x_3 \vee x_4) = 1] \\
&= 3\epsilon(1 - \epsilon)^2 + \epsilon^2(1 - \epsilon) \\
\\
\delta^*(\epsilon) &\geq \Pr[y \neq (x_1 \vee x_2) \wedge (x_3 \vee x_4) | (x_1 \vee x_2) = 1 \\
&\quad \text{and } (x_3 \vee x_4) = 0, \text{ or } (x_1 \vee x_2) = 0 \text{ and } (x_3 \vee x_4) = 1] \\
&= 2\epsilon(1 - \epsilon)^2 + \epsilon^2(1 - \epsilon) + \epsilon^3 \\
\\
\delta^*(\epsilon) &\geq \Pr[y \neq (x_1 \vee x_2) \wedge (x_3 \vee x_4) | (x_1 \vee x_2) = 0 \\
&\quad \text{and } (x_3 \vee x_4) = 0] \\
&= \epsilon(1 - \epsilon)^2 + 3\epsilon^2(1 - \epsilon).
\end{aligned}$$

Since $\delta^*(\epsilon)$ is the maximum of the previous three bounds, it is easy to see that for $\epsilon \leq 1/2$,

$$\delta^*(\epsilon) = 3\epsilon(1 - \epsilon)^2 + \epsilon^2(1 - \epsilon),$$

which is strictly decreasing on $(a, 1)$, where $a = (5 - \sqrt{7})/6 \approx 0.39$. Intuitively, this happens because in such a circuit, when ϵ increases, it is more likely that the errors occurring at the two gates cancel out each other.

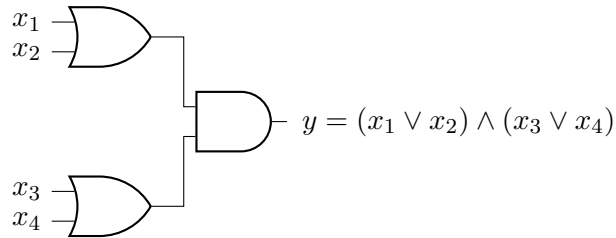


Figure 6: A simple circuit where $\delta^*(\epsilon)$ is not monotone in ϵ in the von Neumann failure model, consisting of two OR gates and one AND gate.

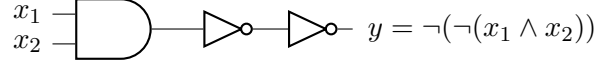


Figure 7: A simple circuit where $\delta^*(\epsilon)$ is not monotone in ϵ in the 0-default failure model, consisting of an AND gate and two NOT gates.

Figure 7 depicts an example where $\delta^*(\epsilon)$ is not monotone in the 0-default failure model. Here there are only two cases to bound $\delta^*(\epsilon)$, since if the inputs are not both 1, the output of the AND gate is a 0 with probability 1.

$$\begin{aligned}
 \delta^*(\epsilon) &\geq \mathbf{Pr}[y \neq x_1 \wedge x_2 | x_1 = 1 \text{ and } x_2 = 1] \\
 &= (1 - (1 - \epsilon)^2)(1 - \epsilon) + \epsilon(1 - \epsilon)^3 \\
 &= 1 - \epsilon - (1 - \epsilon)^4
 \end{aligned}$$

$$\begin{aligned}
 \delta^*(\epsilon) &\geq \mathbf{Pr}[y \neq x_1 \wedge x_2 | x_1 = 0 \text{ or } x_2 = 0] \\
 &= \epsilon
 \end{aligned}$$

Since $\delta^*(\epsilon)$ is the maximum of the previous two bounds, we have that for $\epsilon \leq 0.45$,

$$\delta^*(\epsilon) = 1 - \epsilon - (1 - \epsilon)^4,$$

which is strictly decreasing on $(b, 1)$, where $b = 1 - 4^{-1/3} \approx 0.37$.

6.0 ALMOST ALL FUNCTIONS REQUIRE EXPONENTIAL ENERGY

In this chapter, we show that almost all functions require circuits using exponential energy in the exact failure model. We first show that directly applying Shannon's argument that almost all functions require circuits with exponentially many gates is not sufficient, as in the exact failure model, there are circuits that can compute a logarithmic number of functions using homogeneous voltage supplies, and an exponential number of functions if allowed heterogeneous voltage supplies. We then prove sufficiently small upper bounds on the number of functions a single circuit can compute, both with homogeneous and heterogeneous voltage supplies, and thus show that almost all functions require exponential energy.

6.1 A LOWER BOUND ON THE NUMBER OF FUNCTIONS COMPUTABLE BY A CIRCUIT

In this section we show that in the exact failure model, in both the homogeneous case and the heterogeneous case, a single circuit can reliably compute many different functions. Both of these lower bounds demonstrate that Shannon's counting argument will not be sufficient to show that almost all functions require exponential energy. The lower bound on the number of such functions is much stronger in the heterogeneous case, and thus also demonstrates the power that heterogeneity affords the circuit designer.

6.1.1 Homogeneous Supply Voltages

We start with the homogeneous case giving an explicit construction of a circuit that computes approximately $\log n$ different functions in the exact failure model. The key concept used throughout is that for a large enough perfect binary tree of AND gates (referred to as an AND tree) there is some ϵ such that, regardless of the input, the tree will output 0 with high probability. By combining such trees of different sizes into a single circuit we can essentially ignore different parts of the input depending on ϵ . The statement and proof are formalized below, after the statement of a technical lemma that we need in the proof of our result.

Lemma 52. *Let $\epsilon \leq 1/10$ and $p_0 = 1$, and let $p_i = p_{i-1}^2(1 - 2\epsilon) + \epsilon$. Then, for $i \geq \log(1/\epsilon) + \log \log(1/\epsilon) + 5$, $p_i \leq 3\epsilon$.*

Proof. It is straightforward to show that since $p_0 > \epsilon$, we have $p_i \geq p_{i+1}$ for all i . We first show that $p_{\log(1/\epsilon)+1} \leq 1/\sqrt{2}$. Note that, for $p_{i-1} \geq 1/\sqrt{2}$, we have $p_i = p_{i-1}^2(1 - 2\epsilon) + \epsilon = p_{i-1}^2 - \epsilon(2p_{i-1}^2 - 1) \leq p_{i-1}^2$. Therefore, since $p_1 = (1 - \epsilon)$ it follows that, for $p_i \geq 1/\sqrt{2}$, it holds that $p_{i+1} \leq (1 - \epsilon)^{2^{i-1}}$. Note that for $i = \log(1/\epsilon) + 1$ we have $p_i \leq 1/\sqrt{2}$, as otherwise we would reach a contradiction since we would have $p_i \leq (1 - \epsilon)^{1/\epsilon} \leq 1/e \leq 1/\sqrt{2}$. It is easy to see that $p_{\log(1/\epsilon)+5} \leq 1/8$.

We now show that $p_{\log(1/\epsilon)+5+\log \log(1/\epsilon)} \leq 3\epsilon$. Let $k = \log(1/\epsilon) + 5$. We show that $p_{k+j} \leq p_k^{2^j} + 2\epsilon$ for any $j \geq 0$. To see this note that it clearly holds for $j = 0$, and by induction if this holds for an arbitrary $j > 0$, then

$$p_{k+j+1} = p_{k+j}^2(1 - 2\epsilon) + \epsilon \leq (p_k^{2^j} + 2\epsilon)^2 + \epsilon \leq p_k^{2^{j+1}} + 2\epsilon,$$

since $\epsilon \leq 1/10$ and $p_k \leq 1/8$. By solving $p_k^{2^j} = \epsilon$ we obtain that $p_{k+\log \log(1/\epsilon)} \leq 3\epsilon$. \square

Theorem 53. *In the exact failure model, for any $\delta \in (0, 1/2)$ and $n \in \mathbb{N}$, there exists a circuit C with n inputs and size $O(n)$ that computes $\Omega\left(\frac{\log n}{\log(\frac{1}{\delta} \log n)}\right)$ different functions $(1 - \delta)$ -reliably.*

Proof. The circuit, which we indicate with C , consists of k perfect binary trees of AND gates, which we refer to as $\text{AND}_1, \dots, \text{AND}_k$, and of a complete binary tree of OR gates, denoted OR_1 . The size of AND_i , which will be determined later but decreases exponentially

as i increases, is denoted by s_i , and the size of OR_1 is $k - 1$. Each AND tree receives its own set of input bits. The outputs of these k trees are fed into the tree of OR gates, and the output of the latter tree is the output of the circuit (see Figure 8). Thus, when $\epsilon = 0$, the circuit C computes $\text{OR}(\text{AND}_1, \dots, \text{AND}_k)$.

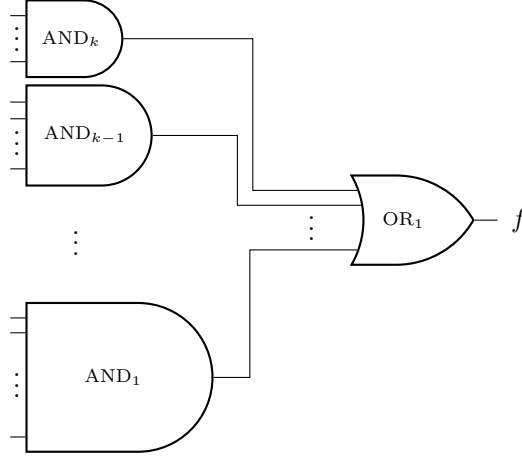


Figure 8: The circuit used in the proof of Theorem 53.

The high level approach is to show that as ϵ grows larger, the larger AND trees switch from computing the AND function to computing the 0 function. In other words, the result is completely determined by the remaining functional AND trees. By choosing the sizes s_i to be sufficiently different, we can show that each AND_i will switch to computing the 0 function at a different ϵ , and further, when this switch occurs all of the smaller trees will still be functioning correctly with high probability.

Before diving into the technical details, we show the two main bounds we need to hold and how with these we can prove the desired theorem. We will use ϵ_i to denote the minimum gate failure probability at which AND_i begins to compute the 0 function with high probability. This value along with the size of each AND_i will be given later. Let $f_i(I) = \text{OR}(\text{AND}_{i+1}(I), \dots, \text{AND}_k(I))$, i.e., the function we are trying to compute. The first result we will need is an upper bound on the probability that the largest AND trees, $\text{AND}_1, \dots, \text{AND}_i$, output a 1. Let E_1^j be the event that AND_j outputs a 1 on any input

when gates are failing with probability ϵ_i . For all j , $1 \leq j \leq i$, we will show that

$$\Pr[E_1^j] \leq \frac{\delta}{2k}. \quad (6.1)$$

The other inequality we will need bounds the probability that any gate in $\text{AND}_{i+1}, \dots, \text{AND}_k$ fails. More specifically, let E_2^j be the event that some gate in AND_j fails. We will show that for all j , $i < j \leq k$,

$$\Pr[E_2^j] \leq \frac{\delta}{2k+2}. \quad (6.2)$$

Let E_3 be the event that any gate in OR_1 fails. We will show that $\Pr[E_3] \leq \Pr[E_2^k]$. The result then follows by applying the union bound over all of the trees and combining the two previous inequalities. That is,

$$\begin{aligned} \Pr[C(I) = f_i(I)] &\geq \Pr[\neg(E_1^1 \vee \dots \vee E_1^i) \wedge \neg(E_2^{i+1} \vee \dots \vee E_2^k \vee E_3)] \\ &\geq (1 - \delta/2)(1 - \delta/2) \\ &> 1 - \delta. \end{aligned}$$

To complete the proof we will show that for specific values of ϵ_i and s_i , Inequality 6.1 and Inequality 6.2 both hold. Let $s_i = (6k/\delta)^{2(k-i+1)}$ and let $\epsilon_i = (\delta/6k)^{2(k-i)+1}$.¹ Fix i , and consider when each gate in C fails independently with probability ϵ_i . We first prove Inequality 6.2. The basic idea is that $1/\epsilon_i$ is much larger than s_{i+1} , and thus the probability that any gate will fail is quite small. More formally, note that by the union bound, since each gate fails with probability ϵ_i and there are s_j gates,

$$\begin{aligned} \Pr[E_2^j] &\leq \epsilon_i s_j \\ &= \left(\frac{\delta}{6k}\right)^{2(k-i)+1} \left(\frac{\delta}{6k}\right)^{-2(k-j+1)} \\ &\leq \left(\frac{\delta}{6k}\right)^{2(k-i)+1} \left(\frac{\delta}{6k}\right)^{-2(k-(i+1)+1)} \\ &\leq \frac{\delta}{2k+2}. \end{aligned}$$

¹For the sake of simplicity we assume that s_i is a power of two, minus one.

The fact that $\Pr[E_3] \leq \Pr[E_2^k]$ follows from noting that OR_1 has $k - 1$ gates, which is less than $s_k = (6k/\delta)^2$.

The last piece is to show that, for these values of s_i and ϵ_i , Inequality 6.1 holds. The key to this argument revolves around a recurrence that describes the probability a gate at some level ℓ in an AND tree outputs a 1 when gates are failing with probability ϵ_i . Intuitively this probability should be decreasing as we go deeper into the tree, so the goal is to show that the tree is large enough so that the root outputs a 0 with sufficiently high probability. Let p_ℓ be the probability that a gate at level ℓ outputs a 1 (where the leaves are level 0) assuming that all inputs are 1. Then, we have that $p_0 = 1$, and

$$p_\ell = (1 - \epsilon_i)(p_{\ell-1})^2 + 2\epsilon_i(1 - p_{\ell-1})p_{\ell-1} + \epsilon_i(1 - p_{\ell-1})^2 = p_{\ell-1}^2(1 - 2\epsilon_i) + \epsilon_i.$$

By Lemma 52, we have that for $\ell \geq \log(1/\epsilon_i) + \log \log(1/\epsilon_i) + 5$, $p_\ell \leq 3\epsilon_i$. Using this, we are now ready to prove Inequality 6.1. Since the height of AND_j is $\log s_j$, we have, for k larger than some constant,

$$\begin{aligned} \log s_j &\geq \log s_i \\ &= \log \left(\frac{6k}{\delta} \right)^{2(k-i+1)} \\ &\geq \log \left(\frac{6k}{\delta} \right)^{2(k-i)+1} + \log \log \left(\frac{6k}{\delta} \right)^{2(k-i)+1} + 5 \\ &= \log \left(\frac{1}{\epsilon_i} \right) + \log \log \left(\frac{1}{\epsilon_i} \right) + 5 \end{aligned}$$

and thus by Lemma 52, $\Pr[E_1^j] \leq 3\epsilon_i \leq \frac{\delta}{2k}$.

Up until this point we have constructed a circuit that computes k different functions, but we have not yet compared k to n . Note that the size of C , which is simply the sum of the s_i 's along with the size of OR_1 (which is of size $k - 1$), is $\Theta(n)$. Therefore, we have that

$$n = \Theta \left(k - 1 + \sum_{i=1}^k s_i \right) = \Theta \left(k - 1 + \frac{\left(\frac{6k}{\delta} \right)^{2k+2} - \left(\frac{6k}{\delta} \right)^2}{\left(\frac{6k}{\delta} \right)^2 - 1} \right).$$

Thus,

$$\frac{\log n}{\log(\frac{1}{\delta} \log n)} = \Theta \left(\frac{(2k+3) \log \frac{6k}{\delta}}{2 \log(\frac{2k}{\delta} \log \frac{6k}{\delta})} \right) = \Theta(k),$$

and the proof is complete. \square

6.1.2 Heterogeneous Supply Voltages

We now show that with heterogeneous voltage settings in the exact failure model, we can construct a circuit that computes exponentially many functions $(1 - \delta)$ -reliably. We leverage the power of heterogeneity to ensure that certain parts of the circuit compute correctly with high probability, while other parts can fail with high probability. In particular, we build a circuit for a CNF formula where the literals of the formula can be determined dynamically by forcing certain gates to fail while preserving the correctness of the CNF calculation. This allows a single circuit to compute all possible functions representable by CNF formulas with n inputs and a fixed number of fixed-length clauses.

Theorem 54. *In the exact failure model, for any constant $\delta \in (0, 1/2)$ and $n \in \mathbb{N}$, there exists a circuit C with n inputs of size $O(n^2)$ that computes $\Omega(3^n)$ different functions $(1 - \delta)$ -reliably.*

Proof. We give a circuit that computes at least 3^n different functions. We delay the discussion of voltages and correctness until we have completely described the circuit. Consider a 3CNF formula Φ with n variables and m clauses, i.e., $\Phi(x)$ is 1 if x satisfies all the clauses and 0 otherwise. To build a circuit that computes Φ , for each clause $(\ell_1 \vee \ell_2 \vee \ell_3)$ we have a single OR gate the inputs of which are variables ℓ_1, ℓ_2, ℓ_3 (note these need not be different and we are ignoring negations here). The output of each such OR gate is fed into an AND tree which outputs the conjunction of all such clauses. In such, this circuit computes the function

$$f_\Phi(x) = \begin{cases} 1 & \text{if } \Phi(x) = 1, \\ 0 & \text{if } \Phi(x) = 0. \end{cases}$$

We now give the construction of the circuit C . Consider a generic 3CNF formula $\Phi = (\ell_1 \vee \ell_2 \vee \ell_3) \wedge \cdots \wedge (\ell_{3m-2} \vee \ell_{3m-1} \vee \ell_{3m})$, and the corresponding series of OR and AND gates as described above, however with input wires coming into each ℓ_i removed. We will use a selection circuit to dynamically connect each ℓ_i to some x_j depending on voltages.

We define the selection circuit for ℓ_i, S_i as follows. This circuit takes as input $\log 2n$ bits as selectors as well as the $2n$ bits $(x_1, \neg x_1, \dots, x_n, \neg x_n)$. The output of S_i is the bit corresponding to the location determined by the first $\log 2n$ bits. Note that Pippenger

provides such a circuit of size $O(n)$ in [25]. Hence for all possible Φ , by appropriately setting the $\log 2n$ bits of each selection circuit, this circuit computes the function f_Φ .

The last piece necessary to define C is describing how the $\log 2n$ input bit b_k of each selection circuit are set. For each such b_k , we have a tree of AND gates of with $\lceil \log \frac{12m \log 2n}{\delta} \rceil = |I|$ inputs, the output of which is fed into b_k . The input to these AND gates are constant 0's that go through a single NOT gate. For each such selection circuit S_i and each such input bit b_k we refer to this circuit as $I_{i,k}$.

We now have a complete description of C and proceed to proving there exist voltage settings such that this circuit correctly computes 3^n different functions. We break this into two parts. We first show that for any fixed Φ there are voltage settings such that, with probability at least $1 - \delta$, $C(x) = f_\Phi(x)$. We then give a bound on the number of unique functions $f_\Phi(x)$.

Fix Φ , and consider the following voltage setting. For each gate in the 3CNF circuit set $\epsilon_S = 1 - (1 - \delta/2)^{1/3m}$. For each gate in S_i , set $\epsilon_{S_i} = 1 - (1 - \delta/2)^{1/6mN}$ where N is the size of S_i . Finally we need to set the voltages for $I_{i,k}$ for all i and k . Consider some i , $1 \leq i \leq 3m$ and some $1 \leq k \leq \log 2n$. Let x_j be the input (ignoring negations) ℓ_i is to be connected to in Φ and let $b = b_1 b_2 \dots b_{\log 2n}$ be the binary representation of $2j - 1$. There are two cases.

Case 1: $b_k = 0$. In this case, set the ϵ of all NOT gates to $1/2$ and set all other ϵ in $I_{i,k}$ to $1 - (1 - (1/2)^{|I|})^{1/2|I|}$, where $|I|$ is the size of the input to $I_{i,k}$.

Case 2: $b_k = 1$. In this case, set the ϵ of all gates in $I_{i,k}$ to $1 - (1 - \delta/2)^{1/6m \log 2n N}$ where N is the size of $I_{i,j}$.

We are now ready to bound the probability that $C(x) = f_\Phi(x)$. Let E_1 be the event that no gate fails in the 3CNF circuit. Since there at most $3m$ gates in the 3CNF circuit, we have $\Pr[E_1] \geq (1 - \epsilon_S)^{3m} = (1 - \delta/2)$.

Next we bound $\Pr[S_i(x) = x_j]$, where x_j is the input ℓ_i is to connect to in Φ . Again let $b = b_1 b_2 \dots b_{\log 2n}$ be the binary representation of $2j - 1$, and let E_2 be the event that no gate in $S_i(x)$ fails. We have

$$\begin{aligned} \Pr[S_i(x) = x_j] &\geq \Pr[I_{i,1} = b_1 \wedge \dots \wedge I_{i,\log 2n} = b_{\log 2n} \wedge E_2] \\ &= \Pr[I_{i,1} = b_1] \cdot \dots \cdot \Pr[I_{i,\log 2n} = b_{\log 2n}] \cdot \Pr[E_2]. \end{aligned}$$

By construction we have $\Pr[E_2] \geq (1 - \epsilon_{S_i})^{|S_i|} = (1 - \delta/2)^{1/6m}$. To show a similar bound for the product of the first $\log 2n$ terms we have two cases. Let $c = 6m \log 2n$.

Case 1: $b_k = 0$. Let E_3 be the event that all NOT gates do not fail and E_4 be the event that no gate in the AND tree of $I_{i,k}$ fails. Then we have that, when n and m are larger than some constant over δ ,

$$\begin{aligned}
\Pr[I_{i,k} = 0] &= 1 - \Pr[I_{i,k} = 1] \\
&= 1 - (\Pr[I_{i,k} = 1|E_3] \Pr[E_3] + \Pr[I_{i,k} = 1|\neg E_3] \Pr[\neg E_3]) \\
&\geq 1 - (\Pr[E_3] + \Pr[I_{i,k} = 1|\neg E_3]) \\
&\geq 1 - ((1/2)^{|I|} + (1 - \Pr[I_{i,k} = 0|\neg E_3])) \\
&\geq 1 - ((1/2)^{|I|} + (1 - \Pr[E_4])) \\
&\geq 1 - (1/2)^{|I|-1} \\
&= 1 - (1/2)^{\log(2c/\delta)-1} \\
&\geq 1 - (1/2)^{\log(c\delta)+8/\delta c} \\
&\geq 1 - (1/2)^{-\log(1/c\delta)-\log(1-4/\delta c)} \\
&\geq 1 - (1/2)^{-\log(1/c\delta-4/c^2\delta^2)} \\
&\geq 1 - (1/2)^{-\log(1-e^{-2/\delta c})} \\
&\geq 1 - (1/2)^{-\log(1-(1-\delta/2)^{1/c})} \\
&= (1 - \delta/2)^{1/c}.
\end{aligned}$$

Some of the above inequalities follow from the Taylor series expansions of e^x and $\log(1-x)$.

Case 2: $b_k = 1$. Note that in this case, $\Pr[I_{i,k} = 1] \geq \Pr[\text{No gate in } I_{i,k} \text{ fails}] = (1 - \delta/2)^{1/6m \log 2n} = (1 - \delta/2)^{1/c}$.

Combining these, we have that $\Pr[S_i(x) = \ell_i] \geq (1 - \delta/2)^{1/3m}$. We are now ready to bound the probability that C correctly computes $f_\Phi(x)$.

$$\begin{aligned}
\Pr[C(x) = f_\Phi(x)] &\geq \Pr[S_1(x) = \ell_1 \wedge \cdots \wedge S_{3m}(x) = \ell_{3m} \wedge E_1] \\
&= \Pr[S_i(x) = \ell_i]^{3m} \cdot \Pr[E_1] \\
&\geq (1 - \delta/2)(1 - \delta/2) \\
&> 1 - \delta.
\end{aligned}$$

Consider the case where $m = n$. We now compute the size of C . The size of the 3CNF circuit is at most $3n$. For each of the $3n$ literals, there is a circuit of size $O(n)$ that uses $\log 2n$ bits to map an input or its negation to that literal. Each of the $O(n \log n)$ bits is created by a tree of size $O(\log(n \log(n)/\delta))$. Thus C has size $O(n^2 + n \log(n) \log(1/\delta))$.

The last step is to show that there are $\Omega(3^n)$ unique functions $f_\Phi(x)$ with m clauses. Consider some subset $S = \{s_1, \dots, s_{|S|}\} \subseteq [n]$ and some setting $x = (x_{s_1}, x_{s_2}, \dots, x_{s_{|S|}})$ for the variables x_i such that $i \in S$. Then, for each such x_i , if $x_i = 1$ create the clause $(x_i \vee x_i \vee x_i)$ and if $x_j = 0$ create the clause $(\neg x_j \vee \neg x_j \vee \neg x_j)$. Create $n - |S|$ additional clauses that are a duplicate of one of these clauses. Note that the resulting formula Φ returns 1 exactly when the input bits S are set to x , regardless of the value of the rest of the input bits, and 0 otherwise. Thus for each unique setting of x and each unique S we obtain a new function. Since there are $\binom{n}{|S|}$ ways to choose S and $2^{|S|}$ settings of x , by summing over $0 \leq |S| \leq n$ we get the desired result. \square

6.2 ALMOST ALL FUNCTIONS REQUIRE EXPONENTIAL ENERGY

In this section we show that, despite the ability of a single circuit to compute multiple functions in the exact failure model, an upper bound on the number of such functions and an adaptation of Shannon's argument allows us to show that almost all functions require exponential energy, both in the homogeneous and heterogeneous case. In some sense, this is evidence that the advantages heterogeneity provides are somewhat limited, as even though some heterogeneous circuits can compute many more functions than any homogeneous circuit

of the same size, this advantage is not sufficient to reduce the minimal circuit size by more than a constant for almost all functions.

6.2.1 Adaptation of Shannon's Argument

Inspired by Shannon's counting argument that almost all functions require exponentially-sized circuits, we show first that, in circuit models where circuits can compute multiple functions, as long as the number of functions a single circuit can compute is not too many, almost all functions still require exponentially-sized circuits. We will combine this with upper bounds on the number of functions homogeneous and heterogeneous circuits can compute to obtain our main results. Note that the following lemma assumes gates have fan-in at most two, and thus all our results assume gate fan-in is at most two; It is straightforward to generalize this lemma and our results to any setting where the fan-in of gates is a constant.

Lemma 55. *Suppose a circuit of size s can compute at most $f(s)$ functions in some circuit model where gates have fan-in at most two. If there exists some constant $c > 0$ such that $s^{4s} f(s) = o(2^{2^n})$ for $s = 2^n/cn$, then almost all circuits require $\Omega(2^n/n)$ gates in that model.*

Proof. Consider the set of circuits with at most s gates. A standard counting argument shows that any circuit in this set can be represented with $4s \log s$ bits, and therefore there are at most s^{4s} circuits with size at most s . Thus, if for some $c > 0$ and $s = 2^n/cn$ it holds that $s^{4s} f(s) = o(2^{2^n})$, then almost all functions require circuits of size at least $2^n/cn = \Omega(2^n/n)$. \square

6.2.2 Homogeneous Supply Voltages

In this subsection we show that almost all functions require exponential-energy homogeneous circuits in the exact failure model. In some sense, this result is a corollary of the later result that almost all functions require exponential-energy heterogeneous circuits; However, we include this result as it illustrates how homogeneous circuits are simpler than heterogeneous circuits, and we are able to obtain a slightly stronger lower bound on the energy used by almost all functions. Our proof aims to bound the number of functions a circuit of size s

can compute, which is necessary, since, as we showed in the previous section, a single circuit can compute many functions.

Lemma 56. *For any circuit C on n inputs with s gates, and any $\delta > 0$, let \mathcal{F} be the set of all functions f for which there exists some ϵ such that (C, ϵ) is $(1 - \delta)$ -reliable for f . Then, $|\mathcal{F}| \leq s2^n + 1$.*

Proof. Fix some circuit C and input I , and let $C_I(\epsilon)$ be the probability that C outputs a 1 on input I with ϵ -faulty gates. Note that by definition for C to compute some function f with ϵ -faulty gates we must have that for all inputs I , either $C_I(\epsilon) \geq 1 - \delta$ or $C_I(\epsilon) \leq \delta$. Fix some input I and consider how the output of C changes as we vary ϵ . Note that the above observation implies that C will only switch the function it is computing due to input I if $C_I(\epsilon) = 1 - \delta$ and $C_I(\epsilon)$ is decreasing or $C_I(\epsilon) = \delta$ and $C_I(\epsilon)$ is increasing. However note that $C_I(\epsilon)$ is a polynomial in ϵ of degree s ,² and therefore there are at most s such points since between any two of them the function must change at least once from increasing to decreasing or vice versa. This means that each input I can cause C to switch the function it is computing at most s times. Since there are 2^n distinct inputs, this means that C can switch functions at most $s2^n$ times, and therefore it is able to compute at most $s2^n + 1$ different functions. \square

Since $E(\epsilon) = \Omega(1)$ for $\epsilon > 1/2$, we need only show that almost all functions require exponentially many gates in this model to show that almost all functions require exponential energy. However, the following lemma will allow us to strengthen our theorem statement, and will be helpful later to show that heterogeneous circuits can asymptotically save energy over homogeneous circuits.

Lemma 57. *Let C be a homogeneous circuit that is $(1 - \delta)$ -reliable. Then, $\epsilon \leq \delta$.*

Proof. Let f be the function C is trying to compute, and fix some input I . It suffices to show that the output gate, g_o , must fail with probability less than δ . Let p be the probability that

²If we fix which gates fail, then the output of C on I is fixed to either 1 or 0. A fixed set of q gates fail with probability $\epsilon^q(1 - \epsilon)^{s-q}$, a polynomial of degree s in ϵ . $C_I(\epsilon)$ can be viewed as the sum over the sets of gates that, when failing, cause C to output 1 on I , of the probability of that set failing.

g_o receives an input I' such that $g_o(I') = f(I)$. Then, note that

$$\begin{aligned}
\Pr[g_o(I') = f(I)] &= p(1 - \epsilon) + (1 - p)\epsilon \\
&= p(1 - 2\epsilon) + \epsilon \\
&\leq (1 - 2\epsilon) + \epsilon \\
&= 1 - \epsilon.
\end{aligned}$$

Since by hypothesis $\Pr[C(I) = f(I)] \geq 1 - \delta$, it follows that $\epsilon \leq \delta$. \square

With this in hand, we can now prove the desired theorem.

Theorem 58. *For any $\delta \in (0, 1/2)$, almost all Boolean functions on n variables require homogeneous circuits using $\Omega(\log^2(1/\delta)2^n/n)$ energy.*

Proof. From Lemma 56 we know that each circuit of size s computes at most $s2^n + 1$ different functions. We now show that for $s = 2^n/4n$, the quantity $s^{4s}(s2^n + 1)$ is asymptotically smaller than 2^{2^n} , the number of functions on n inputs. Plugging in and simplifying we have

$$\left(\frac{2^n}{4n}\right)^{4\frac{2^n}{4n}} \left(\frac{2^n}{4n}2^n + 1\right) \leq \frac{2^{2^n}}{n^{\frac{2^n}{n}}} 2^{2^n} \ll 2^{2^n}.$$

Hence, Lemma 55 implies that almost all homogeneous circuits require $\Omega(2^n/n)$ gates. By Lemma 57, we have $\epsilon \leq \delta$, so each gate uses at least $E(\delta)$ energy. \square

6.2.3 Heterogeneous Supply Voltages

In this section we show that almost all functions require exponential energy in the exact failure model, even when allowed heterogeneous circuits. The approach is similar to the one for the homogeneous case, however the bound on the number of functions a heterogeneous circuit can compute requires some technical results from semi-algebraic geometry.

Lemma 59. *For any circuit C on n inputs with s gates, and any $\delta > 0$, let \mathcal{F} be the set of all functions f for which there exists some $\bar{\epsilon} \in (0, 1/2)^{|C|}$ such that $(C, \bar{\epsilon})$ is $(1 - \delta)$ -reliable for f . Then, $|\mathcal{F}| \leq (8e2^n)^s$.*

Proof. Let $\mathcal{P} \subset \mathbb{R}[X_1, \dots, X_k]$ be a finite set of p polynomials with degree at most d . A *sign condition* on \mathcal{P} is an element of $\{0, 1, -1\}^p$. The *realization* of the sign condition σ in \mathbb{R}^k is the semi-algebraic set

$$\mathcal{R}(\sigma) = \left\{ x \in \mathbb{R}^k : \bigwedge_{P \in \mathcal{P}} \text{sign}(P(x)) = \sigma(P) \right\}.$$

Let $N(p, d, k)$ be the number of realizable sign conditions, i.e., the cardinality of the set $\{\sigma : \mathcal{R}(\sigma) \neq \emptyset\}$. The following theorem is due to Alon.

Theorem 60 ([2]).

$$N(p, d, k) < \left(\frac{8edp}{k} \right)^k.$$

Let $I \in \{0, 1\}^n$ be some input to C , and let $P_I(\epsilon_1, \dots, \epsilon_s)$ be the probability that C outputs 1 on I , when gate i fails with probability ϵ_i . Observe that $P_I \in \mathbb{R}[\epsilon_1, \dots, \epsilon_s]$ and that P_I has degree at most s , since we can compute P_I by summing over all possible subsets of gates that could fail and cause C to output a 1, of the probability that exactly those gates fail and no others (which is a polynomial in $\epsilon_1, \dots, \epsilon_s$, where each ϵ_i has exponent 1).

Let $\mathcal{P} = \{P_I - (1 - \delta) \mid I \in \{0, 1\}^n\}$. Clearly, the cardinality of \mathcal{P} is at most 2^n . Observe that every different function f that C calculates must correspond to a unique realizable sign condition of \mathcal{P} , in the sense that there is some setting of $\bar{\epsilon} = (\epsilon_1, \dots, \epsilon_s)$ such that

1. $P(\bar{\epsilon}) - (1 - \delta) > 0$ on inputs I such that $f(I) = 1$, and
2. $P(\bar{\epsilon}) - (1 - \delta) < 0$ on inputs I such that $f(I) = 0$ (in fact, we need $P(\bar{\epsilon}) - \delta < 0$, an even stronger condition).

By Theorem 60, the number of realizable sign conditions of \mathcal{P} is at most $(8e2^n)^s$, which is thus an upper bound on the number of different functions C can compute. \square

We can now prove the main theorem of this section.

Theorem 61. *For any $\delta \in (0, 1/2)$, almost all Boolean functions on n variables require heterogeneous circuits using $\Omega(2^n/n)$ energy.*

Proof. From Lemma 59 we know that each circuit of size s computes at most $(8e2^n)^s$ different functions. We now show that for $s = 2^n/8n$, the quantity $s^{4s}(8e2^n)^s$ is asymptotically smaller than 2^{2^n} , the number of functions on n inputs. Plugging in and simplifying we have

$$\left(\frac{2^n}{8n}\right)^{4\frac{2^n}{8n}} (8e2^n)^{\frac{2^n}{8n}} \leq 2^{\frac{2^n}{2}} 2^{\frac{2^n(3+2-12-4\log n)}{8n}} 2^{\frac{2^n}{8}} \leq 2^{\frac{5 \cdot 2^n}{8}} \ll 2^{2^n}$$

Hence, Lemma 55 implies that almost all heterogeneous circuits require $\Omega(2^n/n)$ gates. The theorem follows since $E(\delta) = \Omega(1)$ and E is decreasing on the interval $(0, 1/2)$. \square

6.3 RELATING ENERGY AND THE NUMBER OF NOISY GATES

In this section, we show that the Boolean functions that require exponential energy are exactly the Boolean functions that require exponentially many noisy gates. Before formalizing this notion we introduce some additional notation. For any Boolean function f on n variables and any reliability parameter δ , let $NG(f, \delta)$ denote the minimum size of any (heterogeneous) circuit that $(1 - \delta)$ -reliably computes f , and $\widetilde{NG}(f, \delta)$ denote the minimum size of any homogeneous circuit that $(1 - \delta)$ -reliably computes f . Similarly define $\mathcal{E}(f, \delta)$ to be the minimum energy used by any (heterogeneous) circuit that $(1 - \delta)$ -reliably computes f , and $\widetilde{\mathcal{E}}(f, \delta)$ the minimum energy used by any homogeneous circuit that $(1 - \delta)$ -reliably computes f . We are now ready to state the main result of this section.

Lemma 62. *For all Boolean functions f , and for all $\delta < 1/2$,*

$$E(1/2)NG(f, \delta) \leq \mathcal{E}(f, \delta) \leq \widetilde{\mathcal{E}}(f, \delta) \leq E\left(\frac{\delta}{\widetilde{NG}(f, \delta)}\right) \widetilde{NG}(f, \delta).$$

Proof. First observe that $\mathcal{E}(f, \delta) \leq \widetilde{\mathcal{E}}(f, \delta)$. We now prove the leftmost inequality. Let $(C, \bar{\epsilon})$ be the circuit achieving $\mathcal{E}(f, \delta)$ and note that by definition $\mathcal{E}(f, \delta) = \sum_{g \in C} E(\epsilon_g)$. Since E is decreasing, it follows that $E(\epsilon_g) \geq E(1/2)$ for all $g \in C$. Additionally, by definition, $|C| \geq NG(f, \delta)$, and the result follows.

To show the rightmost inequality, fix some Boolean function f , and some δ . Let C be a circuit of size $s = \widetilde{NG}(f, \delta)$, and ϵ the failure probability, such that (C, ϵ) is $(1 - \delta)$ -reliable

on f . If $\epsilon \geq \delta/s$, we are done, since E is decreasing. Note that for a circuit of size s , if gates fail with probability at most δ/s , then by the union bound, the probability that any gate fails is at most δ . Thus, if $\epsilon < \delta/s$, the probability that any gate fails is at most δ . However, this implies that $(C, \delta/s)$ is $(1 - \delta)$ -reliable on f as well, and thus can use energy $E\left(\frac{\delta}{\widetilde{NG}(f, \delta)}\right) \widetilde{NG}(f, \delta)$. \square

If $E(1/2)$ is $\Omega(1)$, and $E\left(\frac{\delta}{\widetilde{NG}(f, \delta)}\right)$ is bounded above by a polynomial in $\widetilde{NG}(f, \delta)$ and $1/\delta$ (recall that in current CMOS technologies $E(\epsilon) = \Theta(\log^2(1/\epsilon))$), this implies that any function that requires exponential energy requires exponential circuit size and vice versa.

7.0 THE POWER OF HETEROGENEITY TO REDUCE ENERGY

In the previous chapter we studied the power of heterogeneity at a very coarse granularity, asking how many functions require exponential energy, paying little attention to the specific functions themselves. In this chapter we turn our focus to individual functions asking the question, for a given function, how much energy does the best heterogeneous circuit save over the best homogeneous circuit? We are able to show that for a wide class of natural functions, heterogeneity allows us to save $\log n$ energy when δ is polynomial in $1/n$. On the contrary we show that for this same class of functions we cannot hope to do any better. That is, we show an upper bound of $\log n$ on the energy savings possible due to heterogeneity. We conclude the chapter by showing that if we extend the setting to include relations, an energy savings of $\log^2 n$ is possible. This continues the theme of demonstrating that while heterogeneity offers the circuit designer tangible benefits, there are proven limitations to these advantages.

7.1 LOWER BOUND FOR FUNCTIONS

We start by showing a lower bound on the energy savings possible with heterogeneity. In particular, we show that, when δ is a polynomial function of the minimum circuit size s , it is possible to obtain an $\Omega(\log s)$ energy savings using heterogeneous voltages in the bounded failure model. The result is that many natural Boolean functions can be computed with asymptotically less energy using heterogeneous circuits. When both s and the number of non-degenerate inputs (see Definition 65) is $\Theta(n)$, this also holds in the exact failure model.

Theorem 63. *For any function f with minimum circuit size s , for any constant $c > 0$, if $\delta = 1/s^c$, in the bounded failure model every homogeneous circuit for f uses $\Omega(s \log^2 s)$ energy, and there exists a heterogeneous circuit using $O(s \log s)$ energy.*

Proof. The first task is to give a lower bound on the energy used by any homogeneous circuit that $(1 - \delta)$ -reliably computes f . By assumption since s gates are required when there are no failures, and, because the circuit is homogeneous, gates (in particular, the output gate) can fail with probability at most $1/s^c$. Since by Lemma 57 it must be that $\epsilon < \delta$, and $E(1/s^c) = \Theta(\log^2 s)$, we have $\Omega(s \log^2 s)$ energy is required.

The upper bound requires significantly more work, although it is still a somewhat straightforward use of techniques from [17] which proves the following, as part of the proof of Theorem 69:

Lemma 64. *Let the maximum fan-in of any gate be a constant. There is a constant $\epsilon_1 > 0$ and $\theta > 1/2$ such that for any $\epsilon \leq \epsilon_1$, there is a $\rho = \rho(\epsilon) < 1$ such that any gate g of fan-in ℓ can be replaced by a gadget with*

1. k input wires for each input to g ,
2. k output wires, and
3. $\Theta(k)$ gates,

with the property that if, for all i , at least a θ fraction of the i -th set of input wires carries bit b_i , then the probability that fewer than a θ fraction of the output wires carries $g(b_1, \dots, b_\ell)$ is at most ρ^k .

In a manner similar to the proof of Theorem 69, we use Lemma 64 to replace each gate in the original circuit with a gadget whose input and output is $\Theta(\log s)$ wires, and set the failure probability of this section of the circuit to ϵ_1 , with the result that the probability that less than a θ fraction of the wires carry the correct output (i.e., the output if there were no failures) is at most $1/s^{c+2}$. Since the failure rate is set to be constant, the first part of the circuit uses energy $\Theta(s \log s)$. The probability that any gadget's output does not carry at least a θ fraction of the correct bits is at most $1/s^{c+1}$.

At the end of the circuit, we use the standard majority circuit (see Figure 9) of size $\Theta(\log s)$ to obtain the output, and set the failure of this section of the circuit to be $1/s^{c+2}$,

thus this section of the circuit uses energy $\Theta(\log^3 s)$ and the probability that any gate in this section of the circuit fails is at most $1/s^{c+1}$. \square

7.2 UPPER BOUND FOR FUNCTIONS

We now show that for a large class of natural functions this $\log n$ savings is the best we can hope to do, in both the exact and bounded failure models. We start with a definition of non-degenerate input bits and then give the main theorem of this section.

Definition 65 (non-degenerate input bit). *We say a function has a non-degenerate input bit b_i if there exists some input I such that $f(I) \neq f(I^{b_i})$.*

Theorem 66. *Let f be a function with b non-degenerate input bits. Then, for any $\delta \in (0, 1/2)$, any circuit C that $(1 - \delta)$ -reliably computes f requires $\Omega(b \log 1/\delta)$ energy.*

Proof. This proof is quite similar to the proofs of Theorem 1 and Lemma 6 from [6], which in turn use ideas from [18] and [14]. It is produced in full here for completeness.

We start by considering the case where we have homogeneous failures and these failures occur on wires, with the same failure-to-energy function used throughout the paper. We show that every non-degenerate input must have $\log(1/\delta)/\log(1/\epsilon)$ wires emanating from it, even if heterogeneous failures are allowed. This combined with the assumption that there are b non-degenerate inputs, that all gates have constant fan-in, and that $P(1/2) > 0$ yields the desired result. We then show how the result when failures occur on wires implies the result when failures occur on gates.

Consider some non-degenerate input bit b_i and let z be some input I such that $f(z) \neq f(z^i)$. Let B be the set of all wires originating from b_i and $|B| = m$ be the number of such wires. For all $\beta \subseteq B$, let $H(\beta)$ be the event that all wires in β fail and the other wires of B are correct. Finally, let β_i denote the subset where

$$\max_{\beta \subseteq B} \mathbf{Pr} [C(z^i) = f(z^i) | H(\beta)] .$$

Note that since we assume C is $(1 - \delta)$ -reliable we must have $\Pr[C(z^i) = f(z_i)] \geq 1 - \delta$ and therefore $\Pr[C(z^i) = f(z^i)|H(\beta_i)]$. Let H_i be the event corresponding to $H(B \setminus \beta_i)$. So for example if β_i is the empty set then H_i is the event where all wires fail. Also, note that since b_i is non-degenerate,

$$\Pr[C(z) \neq f(z)|H_i] = \Pr[C(z^i) \neq f(z^i)|H(\beta_i)] \geq 1 - \delta.$$

Finally since $\delta \geq \Pr[C(z) \neq f(z)] \geq \Pr[C(z) \neq f(z)|H_i] \Pr[H_i] \geq (1 - \delta) \Pr[H_i]$ we get that $\Pr[H_i] \leq \delta/(1 - \delta)$. Combining this with the fact that $\Pr[H_i] \geq \epsilon^m$ and using simple algebra yields $m \geq \log(1/\delta)/\log(1/\epsilon)$ as desired.

Note that in the heterogeneous case when these wires fail with probabilities $\epsilon_1, \dots, \epsilon_m$, for the purposes of this lower bound we have the following optimization problem: We want to minimize $\sum_{i=1}^m P(1/\epsilon_i)$ subject to the constraint that $\prod_{i=1}^m \epsilon_i \leq \delta/(1 - \delta)$. Since for any $\epsilon_1, \epsilon_2 \in (0, 1/2)$, it holds that

$$P(1/\epsilon_1) + P(1/\epsilon_2) \geq 2P(1/\sqrt{\epsilon_1\epsilon_2}),$$

this will be minimized when all ϵ_i are equal. Therefore, for the purposes of this lower bound, we can assume wire failures are homogeneous.

We have that $m \geq \log(1/\delta)/\log(1/\epsilon)$, which implies the energy used by wires coming from each non-degenerate input bit is at least

$$P(\epsilon) \log(1/\delta)/\log(1/\epsilon) = \Omega(\log(1/\delta)),$$

and thus the energy used by the circuit is at least $\Omega(b \log(1/\delta))$.

We now show why the energy lower bound when failures occur on wires implies the energy lower bound when failures occur at gates. Consider any circuit C that $(1 - \delta)$ -reliably computes f . We construct a new circuit C' for computing f , which is identical to C except that both wires and gates may fail, wires of C' incur some non-zero energy consumption (as a function of their probability of failure), and the gates in C' do not consume energy. First we argue that this can be done such that C' is $(1 - \delta)$ -reliable. The following statement was proved in [14].

Statement 67 ([14]). *Let g be a gate with fan-in n_g , in a circuit C where both gates and wires may fail. Furthermore, let $\epsilon \in (0, 1/2)$, $\zeta_g \in [0, \epsilon/n_g]$ and let $g(t)$ be the output of gate g assuming that its input-wires receive input t , and both g and g 's input-wires are faultless. Then there exists a unique value $\eta_g(y, \zeta_g) \in [0, 1]$ such that if*

- *the input wires of g fail independently with probability ζ_g , and*
- *gate g fails with probability $\eta_g(y, \zeta_g)$ when the gate receives input y ,*

then the probability that g does not output $g(t)$ is equal to ϵ .

Observe that if for each wire i entering gate g we set its probability of failure to $\zeta_g = \epsilon/n_g$, we can apply Statement 67 and set the failure probability on gate g when receiving input y to $\eta_g(y, \zeta_g)$. The result is that when the input wires of gate g in C' receive input t , the probability that g does not output $g(t)$ is ϵ_g (the same as the probability of failure of g in the original circuit C). Thus by setting these failure probabilities for each gate and wire in C' we have that, for any input x , C and C' output $f(x)$ with the same probability, and so C' is $(1 - \delta)$ -reliable.

Since the fan-in of any gate is a constant, for any gate $n_g P(\epsilon/n_g) = \Theta(P(\epsilon))$. Thus the energy used by C' is within a constant of the energy used by C . \square

7.3 LOWER BOUND FOR RELATIONS

In this section we prove that, in contrast with the previous section, there are relations where heterogeneous circuits can obtain a $\omega(\log n)$ energy savings over homogeneous circuits. In fact, we show that a natural supermajority relation obtains a $\Theta(\log^2 n)$ energy savings in both the bounded and exact failure models, which is asymptotically the maximum possible savings for any relation that does not require circuits of superlinear size. Formally, we have the following theorem.

Theorem 68. *Suppose $\delta = 1/n^c$ for some constant $c > 0$. Then there is a relation that can be computed by a heterogeneous circuit using $O(n)$ energy, but for homogeneous circuits requires $\Omega(n \log^2 n)$ energy.*

We cite the following general theorem proved by Pippenger in [25] and formalized by Gacs in [17] that will be useful in our construction in this section of the paper.

Theorem 69. *There is an $\epsilon_0 > 0$ such that for any $\epsilon < \epsilon_0$, $\delta \geq 3\epsilon$, and any function f computable by a (faultless) circuit of size s , there is an $(1 - \delta)$ -reliable circuit computing f of size $O(s \log(s/\delta))$ when gates fail with probability at most ϵ .*

The following relation is quite natural. The relation outputs the majority if at least $3/4$ of the bits are the majority, and otherwise we do not care about the output.

Definition 70. *The Supermajority Relation (SR) is the following Boolean relation:*

$$SR(x) = \begin{cases} 0, & \text{if the number of 0's in } x \text{ is at least } 3n/4, \\ 1, & \text{if the number of 1's in } x \text{ is at least } 3n/4, \text{ and} \\ 0 \text{ and } 1 & \text{otherwise,} \end{cases}$$

where x is the input and $|x| = n$.

Lemma 71. *When $\delta = 1/n^c$, for some constant $c > 0$, SR can be computed by a circuit with heterogeneous voltages using $O(n)$ energy.*

Proof. For simplicity, we assume $n = 2^k - 1$ for some positive integer k . Throughout this proof, we consider only the case when the input has $m \geq 3n/4$ 1's, as the case when it has at least $3n/4$ 0's is symmetric. Let $\epsilon = \min\{\epsilon_0, 1/488\}$. The high level idea is that, in a way similar to Pippenger's technique in Theorem 4.1 of [25], we add increasing redundancy to a standard majority circuit so that failures become increasingly rare as we traverse down the circuit, and close to the end of the circuit we switch to the standard majority circuit, and set the failure rate to be sufficiently low. The majority circuit we modify is a tree of 1-bit full adders; see Figure 9. The majority circuit is composed of $\log n$ levels, where level 1 is the level that takes the input bits as input. When $n = 2^k - 1$ for some positive integer k , then $y_{\log n}$ is the output bit of the circuit (in general, the output will be some function of $y_1, \dots, y_{\log n}$ that is computable by a circuit of size $o(n)$). As an adder is composed of five gates (see Figure 10), for simplicity we think of an adder as a component that fails with probability at most 5ϵ (i.e., it fails when at least one of its gates fails).

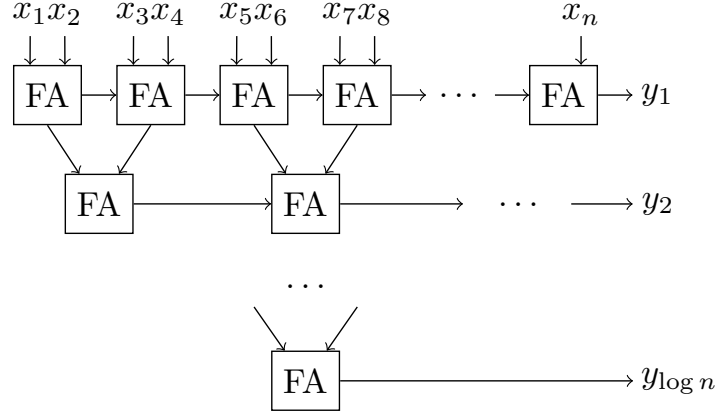


Figure 9: A tree of adders. The majority of x_1, \dots, x_n is a function of $y_1, \dots, y_{\log n}$ that is computable by a circuit of size $o(n)$.

We now describe the modified circuit, which consists of two distinct parts: One for levels at most $\log(n)/6$, and the other for the remainder of the circuit.¹ We describe and analyze the first part here, and describe and analyze the second part after that. The condition needed by the second part of the circuit is that, with probability at least $1/2n^c$, the majority of adder modules on level $\log(n)/6$ output a majority of 1's on their carry wires.

We replace each adder on level $k \leq \log(n)/5$ with a level k adder module. This module has $2k - 1$ inputs for each wire in the original circuit coming from the previous level, as well as $2k + 1$ inputs from the adder to its left's sum bit, and outputs $2k + 1$ wires for both the sum bit and carry bit output. The adder module consists of $2k + 1$ copies of the following: To an adder, supply the fault tolerant majority of the $2k - 1$ wires for each of the two inputs from the previous level, and the fault tolerant majority of the $2k + 1$ wires of the sum bit input, and output the sum and carry bits. We say that an adder module fails if the majority of its output wires do not contain the correct sum and carry bits based on the majority of its input wires. Since each majority circuit can be done with $O(k)$ gates

¹Another possible construction is to replace *all* adders with modules, and add a majority circuit at the end that fails with very low probability. Although this alternate circuit construction could be considered simpler, the analysis appears slightly more complicated.

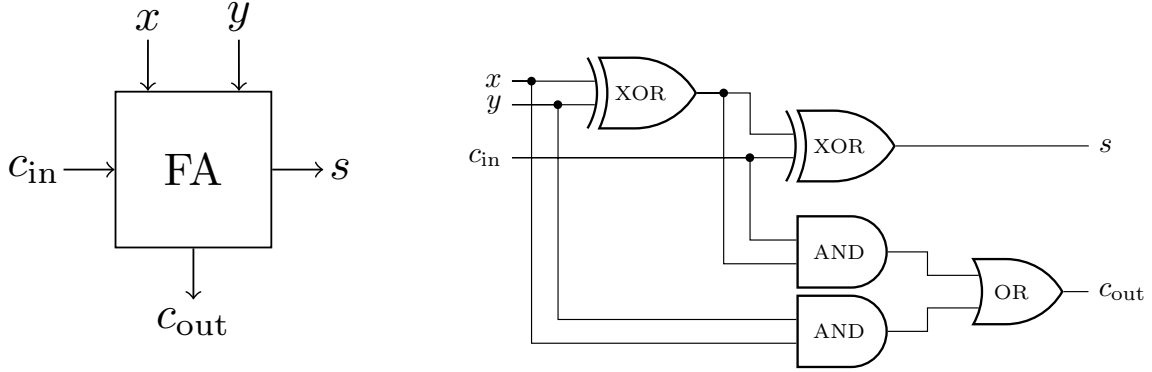


Figure 10: A 1-bit full adder: logic block (left) and circuit realization (right).

without failures, by Theorem 69 there exist majority circuits of size $O(k \log(k/\epsilon))$ that are incorrect with probability at most 3ϵ gates when gates fail with probability ϵ . Since the adder fails with probability at most 5ϵ , the probability of failure of an adder module is at most $p_k = 2^{2k+1}(14\epsilon)^{k+1} = (56\epsilon)^{k+1}/2 < 1/2^{3k+1}$, by our choice of ϵ .

Each adder module thus consists of $O(k^2 \log(k/\epsilon))$ gates, and so the total number of gates in the first part of the circuit is

$$O\left(\sum_{k=1}^{\log(n)/5} \frac{n}{2^k} k^2 \log(k/\epsilon)\right) = O(n),$$

since ϵ is a constant. This also implies that the energy used by the first part of the circuit is $O(n)$, since ϵ is a constant.

It remains to show that the condition needed for the second part of the circuit holds, namely that, with probability at least $1/2n$, the majority of adder modules on level $\log(n)/5$ output a majority of 1's on their carry wires. Let m_k be the number of adder modules at level k that output a majority of 1's on their carry wires. We show that at level k , with high probability

$$m_k \geq \frac{n}{2^k} \left(\frac{3}{4} - \frac{1}{4} \sum_{i=1}^k \frac{1}{2^i} \right).$$

We have the following observation.

Observation 72. *For any fixed input to level k , suppose some subset S of the adder modules fail in any arbitrary way. Let \tilde{m}_k be the number of adder modules at level k that output a majority of 1's on their carry wires, if every module in S failed such that both the sum wires and the carry wires had a majority of 0's Then*

1. $\tilde{m}_k \leq m_k$.
2. $\tilde{m}_k \geq \lfloor (m_{k-1} - 3|S|)/2 \rfloor$.

Suppose there are ℓ_k failures on level k . The above observation allows us to conclude that $m_k \geq \lfloor (m_{k-1} - 3\ell_k)/2 \rfloor$, and so the result follows by induction if we can show $\ell_k \leq \frac{n}{2^k} \frac{1}{3 \cdot 2^{k+1}} - 1$ with sufficiently high probability.

The base case, level 0 (i.e., the carry wires are the input bits), is obvious. For the inductive step, we will use a Chernoff bound to show that ℓ is sufficiently small with high probability. The expected number of adder modules to fail is at most $\mu_k \leq p_k n / 2^k \leq n / 2^{4k+1}$. By a standard Chernoff bound, since modules fail independently, the probability that $n / 2^{4k} \leq \frac{n}{2^k} \frac{1}{3 \cdot 2^{k+1}} - 1$ modules fail is at most

$$\Pr[\ell_k \geq 2\mu_k] \leq \exp\left(-\frac{1}{3} \frac{n}{2^{4k+1}}\right) \leq \exp\left(-\frac{n^{1/5}}{6}\right)$$

since $k \leq \log(n)/5$. By the union bound, the probability that $m_{\log(n)/5} < n^{4/5}/2$ is at most $\log(n) \exp\left(-\frac{n^{1/5}}{6}\right) / 5 < 1/2n^c$ for n large enough.

It remains to describe and analyze the second part of the circuit. From the first part, we receive $n^{4/5}$ sets of $2\log(n)/5 + 1$ wires, and, with probability at least $1/2n^c$, at least half of which carry a majority of 1's. The main idea is to set the failure rate low enough so that no gate fails with high enough probability, and take the majority of each set of input wires, and compute the majority of the values obtained. For each set of wires, we first compute the majority of them, using the standard majority circuit. Since there are $2\log(n)/5 + 1$ wires, this can be done with a circuit of size $O(\log n)$, so in total these majority circuits use $O(n^{4/5} \log n)$ gates. Let $z_1, \dots, z_{n^{4/5}}$ be the results of this. We then use the standard majority circuit to compute the majority of $z_1, \dots, z_{n^{4/5}}$, which uses $O(n^{4/5})$ gates. Thus the total gates in this part of the circuit is $O(n^{4/5} \log n) = o(n)$. We set the failure rate in this part of the circuit to be $1/2n^{c+2}$, so since there are only $o(n)$ gates, the probability

that even one gate fails is at most $1/2n^c$. The energy used by this part of the circuit is $O(n^{4/5} \log^3 n) = o(n)$. \square

We can now prove our main theorem, which is straightforward given the previous lemma.

Proof of Theorem 68. By Lemma 71, SR can be computed by a heterogeneous circuit that uses $O(n)$ energy. It remains to show that any homogeneous circuit computing SR uses $\Omega(n \log^2 n)$ energy. Note that by Lemma 57, gates in any homogeneous circuit computing SR cannot fail with probability more than δ , and since $\delta = 1/n^c$, the energy used by each gate must be at least $\Omega(\log^2 n)$. Additionally, it is obvious that any circuit correctly computing SR must have gates connected to at least half the inputs, and so any circuit computing SR using gates of constant fan-in must have $\Omega(n)$ gates. Therefore, any homogeneous circuit computing SR must use $\Omega(n \log^2 n)$ energy. \square

7.4 GENERALIZING THE FAILURE-TO-ENERGY FUNCTION

Throughout this dissertation we have assumed the failure-to-energy function is $E(\epsilon) = \Theta(\log^2(1/\epsilon))$, based on what appears to be the case in the current technology. However, since this may not be the exact function, or the technology may change, it is important to note that our results hold for more general classes of failure-to-energy functions. In this section, for each result in this chapter, we describe the classes of failure-to-energy functions for which the result holds as it is written in the paper (it is also likely that some results, using different proofs, hold for larger classes of functions than those described here). We begin with a definition of one general class of failure-to-energy functions.

Definition 73. *A failure-to-energy function E is called non-vanishing if it is nonincreasing and $\lim_{\epsilon \rightarrow 1/2^-} E(\epsilon) > 0$.*

For Theorem 63, in order for the given heterogeneous construction to save energy over the given homogeneous lower bound, we need E to be non-vanishing, $E(\epsilon) = \omega(\log(1/\epsilon))$, and for some constant $c > 0$, $\log(n)E(1/n^{c+2}) = o(nE(1/n^c))$.

For Theorem 66, we require E to be non-vanishing, $E(\epsilon) = \Omega(\log(1/\epsilon))$, and $E(\epsilon_1) + E(\epsilon_2) \geq 2E(\sqrt{\epsilon_1\epsilon_2})$.

For Theorem 68, in order for the given heterogeneous construction to save the most possible over the given homogeneous lower bound, we need E to be non-vanishing, and for some constant $c > 0$, $E(1/2n^{c+2}) = O(n^{1/5}/\log(n))$.

8.0 CONCLUSION

We have initiated the theoretical study of energy-efficient circuits, and provided the first results in this area. Leveraging the previous work on fault-tolerant circuit design, we first showed general upper bounds, in terms of circuit size, and lower bounds, in terms of sensitivity, on the energy required to compute a function, when the reliability parameter δ is a fixed constant. Using these results, we showed that when δ is a fixed constant, there is a natural class of functions that do not obtain asymptotically more energy savings when supply voltages may be heterogeneous over the optimal circuit when supply voltages must be homogeneous, indicating that in this setting, allowing heterogeneous supply voltages does not always asymptotically save energy over homogeneous supply voltages. However, we showed that for a specific supermajority relation, and a natural circuit for that relation, a heterogeneous setting of the supply voltages does yield an asymptotic decrease over the energy used by the best homogeneous voltage setting of that circuit, indicating that for fixed, natural circuits, heterogeneous supply voltages may provide an energy savings.

We also considered the complexity of minimizing the energy used by a fixed circuit. We showed that the traditional approach of increasing the voltage of a circuit such that no gate in the circuit fails with sufficient probability is a $\log^2 n$ approximation algorithm to this problem. We also showed that it is NP-Hard to approximate the minimum energy of a circuit to within a factor significantly less than this, indicating that there is a complexity-theoretic barrier to reducing circuit energy beyond the traditional approach in general. We additionally proved this hardness in a second failure model, indicating that these results are not model specific. We showed that for tree circuits, it is possible to determine in polynomial time the probability that a circuit will output correctly, providing some evidence that it may be possible to bypass these hardness results by considering specific families of circuits.

Our next results considered the amount of energy required to compute an average Boolean function. In the bounded failure model, it is straightforward that almost all functions require exponential energy, as this is simply a corollary of Shannon’s classic result that almost all functions require exponential circuit size. In the exact failure model, we found that a single circuit with homogeneous supply voltages can compute a logarithmic number of functions, and a single circuit with heterogeneous supply voltages can compute an exponential number of functions, showing that, in this model, directly applying Shannon’s technique will not work as a single circuit no longer computes a single function. Despite this, we were able to show a sufficiently small upper bound on the number of functions a single circuit can compute using both homogeneous and heterogeneous supply voltages. With this in hand, we showed that almost all functions require exponentially many gates, and thus almost all functions require exponential energy.

We also considered the minimum energy to compute functions and relations when δ must vanish as the number of inputs to the function to be computed increases. We showed that the minimum energy required to compute many functions is a factor of $\log n$ less when heterogeneous supply voltages are allowed, and thus heterogeneous supply voltages can provide significant energy savings over homogeneous supply voltages. We then show that this energy savings is tight for functions with small circuits that do not have degenerate input bits. We additionally showed that a natural supermajority relation can bypass this bound, i.e., the minimum energy circuit with heterogeneous supply voltages is $\log^2 n$ less than the minimum energy circuit with homogeneous supply voltages, and thus relations may potentially obtain greater energy savings via heterogeneous supply voltages than functions.

8.1 OPEN PROBLEMS

There are a number of different, interesting research lines in this area. Here we present those we think are most interesting.

8.1.1 Solving the Minimum Circuit Energy Problem for Restricted Classes of Circuits

Recall that in Chapter 5, we showed that for an arbitrary circuit, it is NP-Hard to approximate the solution to MCE within a factor of $\log^{2-\gamma} n$ for any $\gamma > 0$. However, we were able to give a polynomial time algorithm for determining whether or not a tree circuit is (ϵ, δ) -reliable. This seems to indicate that solving or approximating MCE may be tractable on subclasses of circuits.

A starting point for this would be tree circuits. The main challenge for tree circuits is that, as we showed in Chapter 5, the probability that a circuit is correct does not monotonically decrease with ϵ , and thus one cannot simply binary search over ϵ and use the algorithm we provided for determining if the circuit is (ϵ, δ) -correct. Understanding the relationship between ϵ and the extreme points of the function mapping ϵ to circuit reliability seems to be the key to solving MCE on trees. Trees seem like the easiest case, and are thus a natural starting point. Other, less restrictive, classes of underlying graphs may also make MCE easier, for example circuits whose underlying graph is series-parallel, or has bounded treewidth.

8.1.2 Whether Heterogeneity Reduces Energy When δ is a Fixed Constant

We showed in Chapter 4 that when δ is a fixed constant, some functions, in particular the parity function, and in fact any function with sensitivity $\Theta(n)$ that has circuits of size $\Theta(n)$, do not benefit from heterogeneous supply voltages by more than a constant. It remains open whether or not there exists any function or relation that obtains a non-constant benefit from heterogeneous supply voltages when δ is a fixed constant.

If heterogeneous supply voltages can provide a super-constant energy savings when δ is a constant, it is clear that our techniques from Chapter 7 will not apply, as the energy savings obtained there was in terms of the the energy required for a gate to fail with probability at most δ , which is $\Theta(1)$ if δ is $\Theta(1)$. On the other hand, if reduced energy cannot be obtained via heterogeneous supply voltages, the lower bound techniques from Chapter 3 do not seem very promising, as there does not seem to be a way to apply them to obtain a lower bound

greater than $O(n \log n)$, and thus they could not be used on functions that require circuits of polynomial size in general. The most promising direction to solving this problem seems to be local replacement, i.e., take an arbitrary heterogeneous circuit computing a function, and replace each gate in the circuit with a gadget using homogeneous supply voltages, such that the result still computes the function with only constant increase in energy.

8.1.3 Whether $\log n$ Energy Savings via Heterogeneity is the Maximum Possible When δ Vanishes

We showed in Chapter 7 that when δ is polynomial in $1/n$, a $\log n$ energy savings by using circuits with heterogeneous supply voltages is the maximum possible for functions with circuits of size $\Theta(n)$ that have $\Theta(n)$ non-degenerate input bits. Is it possible this $\log n$ energy savings is the maximum possible for all functions? It may be that the answer to the previously stated open questions, when δ is a fixed constant, provides enough insight to answer this question.

8.1.4 The Power of the Exact Failure Model

One of the difficulties in proving our main result in Chapter 6 that almost all functions require exponential energy was that, in the exact failure model, a single circuit could compute multiple functions. In fact, we were able to show circuits that could compute a logarithmic number of functions if allowed homogeneous voltage supplies, and an exponential number of functions if allowed heterogeneous voltage supplies. However, we do not yet have an example where the circuit designer can use the failures to his advantage to compute functions using less energy. More precisely, are there functions that can be computed with less energy, perhaps even asymptotically, in the exact failure model than in the bounded failure model? Proving the existence of such functions would likely have far-reaching effects in theoretical computer science, as determining whether or not randomness can reduce circuit size remains open, and the problem of reducing the energy needed for computation is very related; Thus this problem may be untouchable given the current state of mathematical knowledge. Still, whether or not it is possible to obtain reduced energy usage in the exact failure model would

be quite interesting, so we present two approaches to possibly answering this question.

First, it may be that the exact failure model allows randomization to be introduced into the circuit, thereby allowing the circuit to be smaller. Gaining an asymptotic decrease in circuit size this way seems difficult given the current state of knowledge of circuits, and in particular the fact that proving a superlinear lower bound on the size of circuits for any function is a longstanding open problem. A constant decrease in circuit size, and therefore energy, may be more possible, though it is still likely very difficult.

Another approach to this problem that is subtly different would be to find a function where instead of using randomization to create a smaller circuit, the circuit computes the function when there are no failures, and also when it has a fixed, high failure rate, but if the failure rate is set adversarially, the circuit does not compute the function (i.e., there is some “intermediate” setting of the voltages that causes the circuit to fail, due to the non-monotonic relationship between voltage and circuit correctness). It is not clear whether or not this approach is more tractable than the approach above.

BIBLIOGRAPHY

- [1] Jaume Abella, Javier Carretero, Pedro Chaparro, Xavier Vera, and Antonio González. Low vccmin fault-tolerant cache with highly predictable performance. In *Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*, pages 111–121. ACM, 2009.
- [2] Noga Alon. Tools from higher algebra. In R. L. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics*, volume 2, pages 1749–1783. MIT Press, 1995.
- [3] Amin Ansari, Shantanu Gupta, Shuguang Feng, and Scott Mahlke. Zerehcache: Armoring cache architectures in high defect density technologies. In *Microarchitecture, 2009. MICRO-42. 42nd Annual IEEE/ACM International Symposium on*, pages 100–110. IEEE, 2009.
- [4] Gary Anthes. Inexact design: beyond fault-tolerance. *Communications of the ACM*, 56(4):18–20, 2013. ISSN 0001-0782. doi: 10.1145/2436256.2436262. URL <http://doi.acm.org/10.1145/2436256.2436262>.
- [5] Antonios Antoniadis, Neal Barcelo, Michael Nugent, Kirk Pruhs, and Michele Scquizzato. Complexity-theoretic obstacles to achieving energy savings with near-threshold computing. In *5th International Green Computing Conference*. IEEE, 2014.
- [6] Antonios Antoniadis, Neal Barcelo, Michael Nugent, Kirk Pruhs, and Michele Scquizzato. Energy-efficient circuit design. In *Proceedings of the 5th conference on Innovations in Theoretical Computer Science (ITCS)*, pages 303–312. ACM, 2014.
- [7] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.
- [8] Neal Barcelo, Michael Nugent, Kirk Pruhs, and Michele Scquizzato. Almost all functions require exponential energy. In *40th International Symposium on Mathematical Foundations of Computer Science*, 2015, in submission.
- [9] Neal Barcelo, Michael Nugent, Kirk Pruhs, and Michele Scquizzato. The power of heterogeneity in near-threshold computing. In *6th International Green Computing Conference*. IEEE, 2015, to be submitted.

- [10] J.A. Butts and G.S. Sohi. A static power model for architects. In *Proceedings of the 33rd annual ACM/IEEE International Symposium on Microarchitecture (MICRO)*, pages 191–201, 2000.
- [11] B.H. Calhoun and A.P. Chandrakasan. A 256-kb 65-nm sub-threshold SRAM design for ultra-low-voltage operation. *IEEE Journal of Solid-State Circuits*, 42(3):680–688, 2007.
- [12] Leland Chang, Yutaka Nakamura, Robert K Montoye, Jun Sawada, Andrew K Martin, Kiyofumi Kinoshita, Fadi H Gebara, Kanak B Agarwal, Dhruva J Acharyya, Wilfried Haensch, et al. A 5.3 ghz 8t-sram with operation down to 0.41 v in 65nm cmos. In *VLSI Circuits, 2007 IEEE Symposium on*, pages 252–253. IEEE, 2007.
- [13] Zeshan Chishti, Alaa R Alameldeen, Chris Wilkerson, Wei Wu, and Shih-Lien Lu. Improving cache lifetime reliability at ultra-low voltages. In *Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*, pages 89–99. ACM, 2009.
- [14] R. L. Dobrushin and S. I. Ortyukov. Lower bound for the redundancy of self-correcting arrangements of unreliable functional elements. *Problems of Information Transmission*, 13:59–65, 1977.
- [15] R. L. Dobrushin and S. I. Ortyukov. Upper bound for the redundancy of self-correcting arrangements of unreliable functional elements. *Problems of Information Transmission*, 13:203–218, 1977.
- [16] Ronald G. Dreslinski, Michael Wieckowski, David Blaauw, Dennis Sylvester, and Trevor N. Mudge. Near-threshold computing: Reclaiming Moore’s law through energy efficient integrated circuits. *Proceedings of the IEEE*, 98(2):253–266, 2010.
- [17] Péter Gács. *Algorithms in Informatics*, volume 2, chapter Reliable Computation. ELTE Eötvös Kiadó, Budapest, 2005. Electronic version also in English: <http://www.cs.bu.edu/faculty/gacs/papers/iv-eng.pdf>.
- [18] Péter Gács and Anna Gál. Lower bounds for the complexity of reliable boolean circuits with noisy gates. *IEEE Transactions on Information Theory*, 40(2):579–583, 1994.
- [19] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001. ISSN 0004-5411. doi: 10.1145/502090.502098. URL <http://doi.acm.org/10.1145/502090.502098>.
- [20] Walid Ibrahim and Valeriu Beiu. Reliability of NAND-2 CMOS gates from threshold voltage variations. In *Proceedings of the International Conference on Innovations in Information Technology (IIT)*, pages 310–314, 2009.
- [21] Jangwoo Kim, Nikos Hardavellas, Ken Mai, Babak Falsafi, and James Hoe. Multi-bit error tolerant caches using two-dimensional error coding. In *Proceedings of the 40th Annual IEEE/ACM International Symposium on Microarchitecture*, pages 197–209. IEEE Computer Society, 2007.

- [22] Timothy N Miller, Renji Thomas, James Dinan, Bruce Adcock, and Radu Teodorescu. Parichute: Generalized turbocode-based error correction for near-threshold caches. In *Proceedings of the 2010 43rd Annual IEEE/ACM International Symposium on Microarchitecture*, pages 351–362. IEEE Computer Society, 2010.
- [23] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in markov chains. *Foundations and Trends in Theoretical Computer Science*, 1(3), 2005.
- [24] Krishna V. Palem. Energy aware computing through probabilistic switching: A study of limits. *IEEE Trans. Computers*, 54(9):1123–1137, 2005.
- [25] Nicholas Pippenger. On networks of noisy gates. In *Proceedings of the 26th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 30–38, 1985.
- [26] Nicholas Pippenger, George D. Stamoulis, and John N. Tsitsiklis. On a lower bound for the redundancy of reliable networks with noisy gates. *IEEE Transactions on Information Theory*, 37(3):639–643, 1991.
- [27] Rüdiger Reischuk and Bernd Schmeltz. Reliable computation with noisy circuits and decision trees—A general $n \log n$ lower bound. In *Proceedings of the 32nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 602–611, 1991.
- [28] Claude E. Shannon. The synthesis of two-terminal switching circuits. *Bell Systems Technical Journal*, 28:59–98, 1949.
- [29] Leslie G. Valiant. Short monotone formulae for the majority function. *Journal of Algorithms*, 5(3):363–366, 1984.
- [30] John von Neumann. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 329–378. Princeton University Press, 1956.
- [31] Chris Wilkerson, Honglliang Gao, Alaa R Alameldeen, Zeshan Chishti, Muhammad M Khellah, and Shiih-Lien Lu. Trading off cache capacity for reliability to enable low voltage operation. In *Computer Architecture, 2008. ISCA’08. 35th International Symposium on*, pages 203–214. IEEE, 2008.
- [32] Gulay Yalcin, Azam Seyedi, Osman S Unsal, and Adrian Cristal. Flexicache: Highly reliable and low power cache under supply voltage scaling. In *High Performance Computing*, pages 173–190. Springer, 2014.